UNCLASSIFIED

AD NUMBER

AD464023

NEW LIMITATION CHANGE

TO

Approved for public release, distribution unlimited

FROM

Distribution authorized to U.S. Gov't. agencies and their contractors; Administrative/Operational Use; DEC 1964. Other requests shall be referred to Office of Naval Research, One Liberty Center, 875 North Randolph Street, Arlington, VA 22203-1995.

AUTHORITY

onr ltr 9 nov 1977

THIS PAGE IS UNCLASSIFIED

DEFENSE DOCUMENTATION CENTER

FOR

SCIENTIFIC AND TECHNICAL INFORMATION

CAMERON STATION ALEXANDRIA, VIRGINIA

Estimating and Detecting the Outputs of Linear Dynamical Systems

by

C. S. Weaver

December 1964

DDC

JUN 7 1955

TISIA B

Technical Report No. 6302-7

464023

SYSTEMS THEORY LABORATORY

STANFORD ELECTRONICS LABORATORIES

STANFORD UNIVERSITY · STANFORD, CALIFORNIA

CATALOGED BY: DDC
AS AD No.

464023

DDC AVAILABILITY NOTICE

ESTIMATING AND DETECTING
THE OUTPUTS OF LINEAR DYNAMICAL SYSTEMS

by

C. S. Weaver

December 1964

Systems Theory Laboratory

Stanford Electronics Laboratories

Stanford University   Stanford, California

ABSTRACT

This investigation considers three closely related problems: the optimum filtering of stationary or near-stationary random processes with unknown parameters from an infinite parameter set; estimation of the state of a linear discrete dynamical system with nongaussian noisy inputs; and applications of state estimation theory to detection. The form of the optimum filter when the parameters are unknown is found to have weights that are averages of simple functions of the signal and noise spectra averaged over the parameter space. Practical methods for implementation are given. The key problem in nonlinear state-variable estimation is obtaining the joint density of the states and the observations in a convenient form. This problem is solved, and surface searching is used to find the mode. The number of dimensions of the surface is the same as the order of the dynamical system. A new approach to linear state estimation is given; and this theory is applied to the problem of detecting a gaussian signal in gaussian noise. A time-invariant, near-optimum detector of small dimensions is derived.

CONTENTS

APPENDIXES

ILLUSTRATIONS

TABLE

SYMBOLS

a_i value of the ith weight of a tapped delay line

a_{ij} the jth element of $\vec{A}^{(i)}$

$e(i)$ the error in estimating $x_1(i)$

$\overline{e^2}$ steady-state mean-squared error of a Wiener filter

$\overline{e_T^2}$ total mean-squared error of all the G_{o_i}

$\overline{\Delta e^2}$ increase in the mean-squared error of $\vec{B}^{(i)}$ over the minimum mean-squared error

$f_{u_\ell}(t)$ characteristic function of $u_\ell(k)$

m_o order of the matrix Φ

$q_i(r)$ output of G_i at the rth sampling time

$u_\ell(k)$ the ℓth component of $U(k)$

$w_\ell(k)$ the ℓth component of $W(k)$

$x_\ell(k)$ the ℓth component of $X(k)$

$z(k)$ a linear combination of the $u_\ell(k)$

A optimum linear estimator of $x_1(k)$

A_k coefficient of $Y(k)$ in filtering equation

$\vec{A}^{(i)}(j,\ell)$ optimum linear estimator of $x_1(i)$ given $Y(j),\ldots,Y(\ell)$

$\triangle A^{(i)}$	difference in impulse response between $\vec{A}^{(i)}$ and $\vec{B}^{(i)}$
B	system bandwidth
B'	output matrix for the two-system combination
B_k	matrix multiplying $\hat{X}(k-1)$ in the filtering equation giving $\hat{X}(k)$
C_k	matrix multiplying $X(1)$ in the filtering equation giving $\hat{X}(k)$ conditioned on $X(1)$
D'	set of observed data
E_j	matrix multiplying $Y(i+j)$ in the equation giving the near optimum estimate of $X(i)$
$E(k)$	error in estimating $X(k)$
$F(\cdot)$	cumulative distribution function (cdf)
G	transition matrix of the second dynamical system
$G_i(j\omega)$	the ith narrowband filter
$G_{o_i}(j\omega)$	the optimum filter at the output of $G_i(j\omega)$
$\mathcal{G}_k$	matrix multiplying $X(1)$ in the equation for the gradient of the exponent of $P[Y(1),\ldots,Y(k)|X(1)]$
H	dynamical-system output matrix
J	output matrix of the combination of two dynamical systems

J_k — that part of the equation for the gradient of the exponent of $P[Y(1),\ldots,Y(k)\,|\,X(1)]$ that is not multiplied by $X(1)$

K — nonideal potentiometer setting

$K_{U(k)}$ — covariance matrix of $U(k)$

$K_{W(k)}$ — covariance matrix of $W(k)$

$K_{\hat{X}(k)}$ — covariance matrix of $\hat{X}(k)$ conditioned on all data to $Y(k)$

$K_{X(k)}(i-m,i+m)$ — covariance matrix of $\hat{X}(k)$ conditioned on $Y(i-m),\ldots,Y(i+m)$

$K_{Y(k)}(k\,|\,k-1)$ — covariance of $Y(k)$ conditioned on all data to $Y(k-1)$

K_{W_n} — covariance matrix of W_n

K_{Y_n} — covariance matrix of Y_n

$L[Y(1),\ldots,Y(n)]$ — likelihood ratio

$N_o(f_i)$ — noise spectrum out of $G_i(j\omega)$

$N(k)$ — noise input to the two-dynamical-system combination

N_i — noise power at the output of $G_i(j\omega)$

$P[\cdot]$ — probability density

$R_x(\tau)$ — autocorrelation function of $x_1(k)$

$R_w(\tau)$ — autocorrelation function of scalar $W(k)$

$R_{x+w}(iT)$ autocorrelation function of signal and noise

S_i signal power at the output of $G_i(j\omega)$

$S_{ii}(f)$ spectrum of signal and noise

S_{ii}^{+} see Eq. (D.1)

$S_{i\bar{i}}$ see Eq. (D.1)

$S_{nn}(f)$ spectrum of noise

$S_o(f_i)$ signal spectrum at the output of $G_i(j\omega)$

$S_{ss}(f)$ spectrum of signal

$S_x(Z)$ sampled-data signal spectrum

$S_{x+w}(Z)$ sampled-data signal plus noise spectrum

T sampling period

$U(k)$ noisy input vector to the dynamical system

$W(k)$ dynamical-system output noise

W_n vector representing a sequence of scalar $W(k)$, $k = 1,2,\ldots,n$

$X(k)$ dynamical-system state vector at the kth sampling time

$\hat{X}(k)$ estimate of $X(k)$

$Y(k)$	the observed random variable at the kth sampling time
Y_n	vector representing a sequence of scalar $Y(k)$, $k = 1, 2, \ldots, n$
α	parameter set
ϕ_{ij}	element of Φ
γ	input matrix for the two-dynamical-system combination
γ_{ij}	element of Γ
ω_i	a state of nature
Γ	the dynamical-system input matrix
Γ_o	path of integration on unit circle
Δ	input matrix for the second dynamical system
Ξ	see Eq. (6.17)
Φ	dynamical-system transition matrix
$\Omega(k)$	output noise generator of the two-dynamical-system combination

ACKNOWLEDGMENT

Appreciation is expressed for the guidance of Dr. Gene F. Franklin, under whom this research was conducted, and for the many helpful suggestions of Dr. Norman M. Abramson. Special thanks are also due Dr. Rupert G. Miller of the Stanford Statistics Department for several helpful conversations—particularly concerning the proof of the first theorem in Chapter VI.

I. INTRODUCTION

A. OUTLINE OF THE PROBLEM

This investigation concerns the optimal estimation or detection of a sampled, vector-valued stochastic process that may be generated by a noisy discrete, linear, dynamical system. The system inputs are a sequence of independent random variables, i.e., white noise. The system output is corrupted with additive white noise. In the first part of this report the white noise is assumed to be gaussian (linear estimation is optimum); later, nongaussian inputs and output noise are assumed (in general, nonlinear estimation will be optimum). The stochastic processes may or may not be stationary and, for most of the report, the process parameters will be assumed to be known a priori. In Chapter V, however, consideration is given to the important problem of estimating the value of the process when it is stationary or nearly stationary and when the parameters are assumed to come from some infinite set with some a priori distribution.

In this analysis, the word "estimation" will mean either filtering or interpolation. An optimum estimate is defined as one that minimizes a generalized mean-squared-error performance criterion or maximizes a conditional density. Filtering is defined as the estimation of a present value conditioned on all the past data. Frequently, the term "smoothing" is used in place of interpolation or estimation of a past value conditioned on all data to the present.

Engineering examples of the above processes are listed below. An important example is a space vehicle in orbit. The equations of motion when linearized correspond to a linear dynamical system. The atmospheric drag may be represented as a noisy input. Range and velocity of the space vehicle are measured over a noisy radio channel. This channel noise constitutes the output noise. Usually, the noise, as it appears to the velocity- and range-measuring equipment, is nongaussian.

An equally important example of this type of system is a rocket under power. The noise inputs are caused by random variations in motor thrust amplitude and direction. The trajectory information is also transmitted over a noisy radio channel.

An example of a stationary process with unknown parameters from an infinite set is a satellite or space vehicle transmitting on an unknown frequency due to uncertainty in the doppler shift or drift in transmitter frequency. Determining and tracking this frequency are the central problems in space communications. For most types of modulation, techniques similar to those discussed in Chapter V provide by far the best solution known. An example of a near-stationary process is a signal with known parameters and unknown jamming, where the jamming corresponds to an unknown output noise. Use of the moon as a passive reflector for communications from one earth point to another is another example of a near-stationary process. We might also include in this class a tracking antenna system using conical scan. Here, the system gain is directly proportional to the unknown signal strength which slowly varies.

B. PREVIOUS WORK

Kalman and Koepcke in their pioneering work [Ref. 1] have considered the optimal filtering and prediction of sampled gauss-markov stochastic processes when the parameters of the process are known. Rauch [Ref. 2] has extended this analysis to include interpolation when the input noise is gaussian, when there is no output noise, and when the parameters are a sequence of independent random variables with known means and variances. Widrow [Ref. 3] and Gabor et al [Ref. 4] have independently investigated and constructed systems that adapt by using a noise-free sample of the signal. Since in many practical situations the noise-free sample will not be available, this type of adaption was not considered in this investigation.

Magill [Ref. 5] has used the Hilbert space approach (approximately concurrently with this investigation) to simplify the derivation of the filter and interpolation of gauss-markov processes. He has also given the form of the optimal estimate of a gauss-markov process when a finite set of parameters is distributed in general according to some arbitrary density. Cox [Ref. 6] discusses state-variable estimation of nonlinear systems with gaussian inputs. His approach involves a system linearization.

Work in this investigation contains an extension of Magill's adaptive estimation for a finite number of parameters to estimation where the parameters may come from an infinite set. The theory of nonadaptive estimation is extended to include dynamical systems with nongaussian inputs and output noise.

C. OUTLINE OF NEW RESULTS

This investigation gives solutions to three closely related problems:

1. The theory of adaptive estimation is extended to stationary or near-stationary processes with parameters from some infinite parameter set.

2. The theory of nonadaptive estimation is extended to include nonlinear filtering and interpolation of the state variables of a dynamical system excited by nongaussian random inputs.

3. Methods are given for greatly simplifying the optimum detection procedures when the signal can be considered as the output of a linear dynamical system excited by random noise.

Chapter II contains a description of the two main mathematical methods of system description that are used in this report. This chapter also contains a detailed description of the random process.

In Chapter III a new approach to the linear estimation of the state variable of a discrete linear dynamical system is presented which shows the close relationship between state-variable estimation and pattern recognition. The estimation theory in Chapter III was inspired by, and is a straightforward extension of, the pattern-learning theory developed by Abramson and Braverman [Ref. 7]. In both cases, it is desired to learn the conditional mean of a gaussian vector-valued random variable. The equations for the mean and the covariance matrices derived by Abramson and Braverman are almost identical to the filtering equations of Chapter III. The author feels that the new approach is far simpler and gives a greater intuitive insight than other methods that have been suggested earlier.

The chapter also provides necessary background for the three chapters that follow. A new problem is solved in Sec. C: estimation of state variables when the estimates are taken through a second dynamical system (as will very often be the case). Equations derived by Magill may be

used to make these estimates; however, the order of Magill's matrix equations can be twice as great as those presented here. If a large amount of data is processed, the reduction in calculation could be significant.

Chapter IV applies the theory of linear state-variable estimation to the problem of detecting a gaussian signal immersed in additive gaussian noise. It is shown that a long-standing conjecture about the possibility of simplifying the detection procedure is often true. In Kailath's solution [Ref. 8], the optimum detector contains an operator that gives the best estimate of each signal sample value during the detection interval based on all the data observed during the interval. If several thousand data points are observed, a matrix of the same order must be inverted. When the signal can be represented as, or approximated by, the output of a noisy dynamical system, the estimation equations of Chapter III may be applied directly. The matrices to be inverted will be no larger than the order of the dynamical system regardless of the number of data points. Further simplification results if the estimator of a sample value is truncated when the error covariance matrix shows that there will be little reduction in mean-squared error by conditioning on additional data points. A near-optimum time-invariant detector is then shown to exist. The advantages of time-invariant circuitry when analog networks are used cannot be overemphasized. Time-variable analog networks of the complexity required in this problem are extremely difficult and expensive to build. This form of detector is the form most convenient to instrument, using the newly developing and powerful methods of optical data processing.

Chapter V describes optimum estimation when the process is stationary or near stationary and when the process parameters are unknown but may assume any one of an infinite number of possible values. The term "near-stationary" is used rather loosely. In practice, the filter will adapt so quickly that good results may be obtained on processes many persons would call highly nonstationary.

The estimator weights are functions of the parameters, and it is shown that the optimum estimator is formed by taking the expected value of the weights over the parameter space conditioned on the observed values.

It is then shown that the parameters enter into the weights as functions of the signal and noise spectrum in a very simple manner. The optimization procedure thus involves learning these spectral functions conditioned on the data.

It is believed that this device will have numerous applications in the field of space communications. As mentioned in Sec. A, a fundamental problem is the frequency tracking of a narrowband signal. The usual approach is to use frequency modulation or to insert an unmodulated carrier. In either case a phase-locked loop may be locked onto the carrier and used as a frequency reference for a narrowband filter. Frequency modulation very often is not the best way to modulate, nor does the inserted carrier contain information, and thus their use lowers the system signal-to-noise ratio for a fixed transmitter power. In electronic surveillance work, the opponent is hardly ever considerate enough to include a tracking carrier! One of his favorite tactics is to shift his transmitter frequency in a manner unknown to the receiver. Such "carrierless" situations show the filter of Chapter V off to good advantage since, unlike the phase-locked loop, it will automatically center itself about the signal in the form of a narrowband filter.

Chapter VI discusses nonlinear estimation (including the best nonlinear predictor) of state variables of linear systems with nongaussian inputs or output noise. With the exception of the types of distributions, the model of the process is identical to that used for linear estimation. First, there is a proof of the necessary and sufficient conditions for the distribution of the state variables to converge to the gaussian. The estimates found in Chapter VI are either Bayesian or maximum likelihood, and the key problem is finding the conditional density in a convenient form. The Markov property of the state variables is used to simplify this rather complex density and then surface-searching techniques are used to find the mode. An important result is proof that near-optimum (linear or nonlinear) estimates of the state of many dynamical systems do not require conditioning on all available data, but may be made using only a short sequence of observations. The length of this sequence may be related directly to the rate of decay of initial conditions in the dynamical

system. The saving in computation time may be very significant. The dimensions of the surface to be searched are the same as the order of the dynamical system.

Use of this theory is envisioned in a situation such as the one given below. Much of the ballistic missile work at Cape Kennedy is concerned with measurement of missile accuracy. The trouble is that the external ground-based measuring equipment is no more accurate than the missile guidance and thus cannot offer any real check on the trajectory. Any increase in guidance accuracy will completely swamp the measuring equipment. (The seriousness of the problem has prompted the government to issue a large contract for range modification, although it is the opinion of many that the point of diminishing returns in measurement-equipment accuracy has already been passed.) In data reduction the standard procedure is to make a linear least-mean-squared estimate. Since it is known that the statistics are nongaussian, nonlinear estimation may offer a possibility of significant improvement. The asymptotic behavior of the estimator as the output noise decreases is also discussed.

II. STATEMENT OF THE PROBLEM AND MODEL OF THE PROCESS

It is desired to form an estimate of a sampled-data, random-message process corrupted by additive noise. The random message or the additive noise or both may be nongaussian. The observable process (the process from which the estimations are made) is assumed to be sampled, either in scalar or vector form, and for most of this report is assumed to be generated by processes with known statistics. In Chapter V, however, it is assumed that the estimates are made of a process with unknown parameters and that these parameters may come from an infinite set.

A. STATE-VARIABLE AND SAMPLE-VALUE REPRESENTATIONS OF DISCRETE LINEAR SYSTEMS

Two mathematical methods for describing linear discrete dynamical systems are used in this report. The first is called the state-variable representation, and the second is known as the sample-value representation. Since many engineers are familiar with one or the other, but not with both, the methods are discussed briefly in this section.

Consider the transfer function

$$G(Z) = \frac{Z^{-2}}{(1 - aZ^{-1})(1 - bZ^{-1})} = \frac{Y(Z)}{u_2(Z)} \qquad (2.1)$$

where $u_2(Z)$ is a noisy control input and $Y(Z)$ is the output. A block diagram having this transfer function is shown in Fig. 1. The input $u_1(k)$ is a second input of noise alone. Let $x_1(k)$ be the value (or state) of the output of the right-hand delay at the kth sampling instant, and let $x_2(k)$ be the value at the output of the other delay. Then the state or state vector of the system is defined as

$$X(k) = \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} \qquad (2.2)$$

In general, any sampled-data transfer function may be reduced to block diagram form with feedback around delays, and a state vector may be defined with the values of the delay outputs at the sampling instants as vector elements.

SEL-64-131
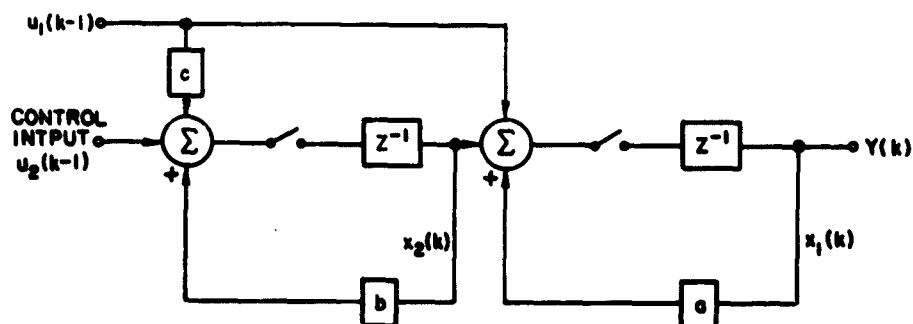
FIG. 1. AN EXAMPLE OF A DISCRETE LINEAR SYSTEM.

The vector $X(k)$ may be found from a set of difference equations that may be written as

$$X(k) = \Phi X(k-1) + \Gamma U(k-1) \tag{2.3a}$$

The matrix Φ is known as a "transition matrix" and is $m_o \times m_o$ where m_o is the order of the system. This matrix may be time variable, but to simplify notation, its argument will not be carried along. The ij^{th} element of Φ, Φ_{ij}, is the gain between the output of the ith delay and the input to the jth delay. The "input matrix" Γ is also $m_o \times m_o$ and it determines where the input vector $U(k-1)$ is applied to the system.

The system output may be a vector or it may be scalar, and it usually is a linear combination of the states. The output can be written as

$$Y(k) = HX(k) \tag{2.3b}$$

where H is a $q \times m_o$ matrix (vector outputs), with q the number of outputs. The H matrix also may be time variable. The system equations for the system of Fig. 1 will then be

$$X(k) = \begin{bmatrix} a & 1 \\ 0 & b \end{bmatrix} \begin{bmatrix} x_1(k-1) \\ x_2(k-1) \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ c & 1 \end{bmatrix} \begin{bmatrix} u_1(k-1) \\ u_2(k-1) \end{bmatrix} \quad (2.4a)$$

$$Y(k) = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} \quad (2.4b)$$

The second mathematical description will be used when the input and the output of the dynamical system are both scalar. The output of a linear system is a linear combination of the inputs, or

$$c(k) = \sum_{j=0}^{k} a_j(k) \, r(k-j) \quad (2.5)$$

where $r(k)$ is the input and $c(k)$ is the output. The system may be thought of as a tapped delay line similar to the one shown in Fig. 12 of Appendix C. If $r(t)$ is a bandlimited continuous function with value $r(k)$ at the kth sampling instant, and if the time between samples, T, is less than $1/2B$, where B is the bandwidth, the well-known sampling theorem [Ref. 9] shows that there is a one-to-one correspondence between the sequence $r(k)$, $k = 0,1,\ldots,k_o$, and $r(t)$ over the interval $[0, k_o T]$. There is also a one-to-one correspondence between the sequence of $c(k)$ and $c(t)$ over the same interval. For a time-invariant system, the $a_j(k)$ are the coefficients in the series expansion of the system Z-transform.

If the $r(k)$, $k = 0,1,\ldots,k_o$ are the (k_o+1) elements of a vector called the "input vector," and if the $c(k)$, $k = 0,1,\ldots,k_o$, are the (k_o+1) elements of the "output vector," the two vectors are related by the following matrix equation. (The square matrix will be called the "transfer matrix.")

$$\begin{bmatrix} c(0) \\ c(1) \\ \vdots \\ c(k_o) \end{bmatrix} = \begin{bmatrix} a_0(0) & & & & \\ a_1(1) & a_0(1) & & & \\ a_2(2) & a_1(2) & a_0(2) & & \\ \vdots & & & \ddots & \\ a_{k_o}(k_o) & & \cdots & & a_0(k_o) \end{bmatrix} \begin{bmatrix} r(0) \\ r(1) \\ \vdots \\ r(k_o) \end{bmatrix} \quad (2.6)$$

SEL-64-131

B. MODEL OF THE PROCESS

The message process will be generated by random inputs, $U(k)$, to the system shown in Fig. 2. It is assumed that $U(k)$ is independent of $U(j)$ for $j \neq k$. The element $u_{\ell}(k)$ is also assumed to be independent of $u_i(k)$ for $\ell \neq i$.
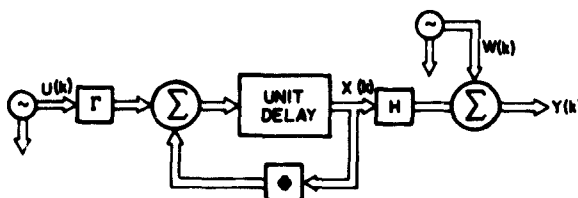
FIG. 2. THE DYNAMICAL SYSTEM.

Much use will be made in this report of a characteristic of $X(k)$ called the (strict) "Markov property." This property is defined by the relationship

$$F[X(k)|X(1),\ldots,X(k-1)] = F[X(k)|X(k-1)] \qquad (2.8)$$

where $F(\cdot)$ is the cumulative distribution function (cdf). In other words, the density of $X(k)$ given all the $X(j)$ up to $X(k-1)$ is the same as the density conditioned only on $X(k-1)$. From Eq. (2.3a) it is seen that if $X(k-1)$ is given, then the only random variable on the right is $U(k-1)$. Since $U(k-1)$ is independent of $U(j)$ for $j \neq k$, the density of $X(k)$ is entirely determined by $X(k-1)$ and $U(k-1)$. If $U(j)$ is a gaussian random variable, the process is known as a "gauss-markov" process.

The message process is assumed to be transferred over a physical channel (such as a telemetering link) and noise will be added. In our model, this noise is indicated by the vector $W(k)$, and Eq. (2.3b) will be modified to

$$Y(k) = HX(k) + W(k) \qquad (2.3c)$$

Estimations of $X(j)$ will be made by observing (the observable process) the sequence $Y(1),\ldots,Y(k)$.

Both $U(k)$ and $W(k)$ are assumed (without loss of generality) to have zero mean. The vectors $U(k)$ and $W(j)$ will be independent for all j and k, and $W(k)$ will be independent of $W(j)$ for $j \neq k$.

Two types of covariance matrices are used frequently in this study. The first is called a "state-variable covariance matrix" and is denoted by the letter K. For example,

$$E[X(k)\ X(k)^t] = K_{X(k)} \tag{2.9a}$$

$$E[Y(k)\ Y(k)^t] = K_{Y(k)} \tag{2.9b}$$

$$E[W(k)\ W(k)^t] = K_{W(k)} \tag{2.9c}$$

where superscript t means the transpose.

The second type of covariance matrix is called the "time-series covariance matrix." An example would be

$$K_S = E \left\{ \begin{bmatrix} s(1) \\ \vdots \\ s(k) \end{bmatrix} [s(1),\ldots,s(k)] \right\} \tag{2.10}$$

III. OPTIMUM LINEAR SMOOTHING AND FILTERING

This chapter discusses linear smoothing and filtering. The estimators will be derived by assuming that the input random process and the output noise are gaussian. So, the first section will contain a brief discussion of linear filtering of nongaussian processes with either gaussian or nongaussian output noise. Section A also contains a comparison of the Bayesian and the maximum likelihood estimation of state variables. Section B contains the derivations of the filtering and smoothing routines. In many practical applications, the observations will be made through a second dynamical system, and Section C contains a derivation of the required modifications of estimating procedures.

A. LINEAR FILTERING

The two classes of estimators to be considered in this report are Bayes estimators and maximum likelihood estimators.

Definition: A Bayes estimator $\hat{X}(j)$ is one that minimizes the expected risk,

$$\rho(\hat{X}) = \int L[\hat{X}(j),X(j)]\ P[X(j)|Y(1),\ldots,Y(k)]\ dX(j)$$

where $L[\hat{X}(j),X(j)]$ is the loss and $\hat{X}(j)$ is the estimate of $X(j)$ given $Y(1),\ldots,Y(k)$.

If

$$L[\hat{X}(j),X(j)] = |\hat{X}(j) - X(j)|^2 \tag{3.1a}$$

it can easily be shown that

$$\hat{X}(j) = E[X(j)|Y(1),\ldots,Y(k)] \tag{3.1b}$$

The loss function for the Bayes estimate will always be that defined in Eq. (3.1a), i.e., minimizing the mean-squared error.

Definition: The maximum likelihood estimate of $X(j)$ is the $X(j)$ that maximizes $P[X(j), Y(1), \ldots, Y(k)]$.

It is frequently convenient to find this maximum by setting the gradient of $\ln\{P[X(j)|Y(1), \ldots, Y(k)]\}$ with respect to $X(j)$ equal to zero. The elements of the gradient vector are known as "likelihood functions."

If $P[X(j)|Y(1), \ldots, Y(k)]$ is symmetric about the highest mode, the conditional mean and the maximum likelihood estimate coincide. Such conditions exist if $X(j)$ and $W(j)$ are gaussian.

In Sec. B, the Bayes estimators for gaussian inputs are seen to be linear. Now, if the actual states are taken from the dynamical system and subtracted from the corresponding estimator output to obtain the error, the system from the noise inputs to the error outputs is still linear. Then, the error covariance matrix will be just a linear transform of the input covariance matrix plus a linear transform of the output noise covariance matrix. The mean-squared error is the trace of the error covariance matrix. In other words, the mean-squared error of the filter that is optimum for gaussian input and output noise is a function only of the input and output noise covariance matrices. If a filter is designed to be optimum for a gaussian input process with a given covariance matrix, the mean-squared error of this filter, when the input is non-gaussian with an identical covariance matrix, will be the same as the gaussian input mean-squared error.

In Chapter IV (for example) it is seen that if one asks the question, What is the linear circuit that will give the minimum mean-squared error? the answer is found to be a function of the input variance and covariance only. Then, the filters derived in this chapter are the best linear ones for all densities with the same covariance matrix. If any improvement is to be gained for nongaussian processes, it necessarily will be a nonlinear operation. Nonlinear smoothing and filtering are discussed in Secs. B and C, Chapter VI.

B. OPTIMUM LINEAR FILTERING AND SMOOTHING

In this section, the optimum filter is first derived. Next, the optimum smoothing routine is found and it will be seen that it contains filtering as a subroutine. The model is shown in Fig. 2 with $W(k)$ and $U(k)$ gaussian. The filtering problem then is to estimate $X(k)$ given all values of Y up to $Y(k)$. As noted in the last section, the optimum estimate is the conditional mean of the quantity to be estimated. In other words, if it is desired to estimate $X(k)$, write the density $P[X(k)|Y(k),Y(k-1),...,Y(1)]$, and find its mean. Later, it will be shown that this mean is identical to the mean of $P[X(k)|Y(k),\hat{X}(k-1)]$, where $\hat{X}(k-1)$ is the estimate of $X(k-1)$ given $Y(1),...,Y(k-1)$. So our problem is reduced to finding this second mean.

Using Bayes' theorem, write

$$P[X(k)|Y(k),\hat{X}(k-1)] = \frac{P[Y(k)|X(k),\hat{X}(k-1)] \; P[X(k)|\hat{X}(k-1)] \; P[\hat{X}(k-1)]}{P[\hat{X}(k-1),Y(k)]} \quad (3.2)$$

Since this density is gaussian, the mean will be the value of $X(k)$ that minimizes the exponent of the density.

Note that $X(k)$ is contained only in $P[Y(k)|X(k),\hat{X}(k-1)]$ and $P[X(k)|\hat{X}(k-1)]$. Referring to Fig. 2, it is seen that

$$P[Y(k)|X(k),\hat{X}(k-1)] = P[Y(k)|X(k)] \approx N[HX(k),K_{W(k)}] \quad (3.3)$$

where $K_{W(k)}$ is the covariance matrix of $W(k)$; [$N(A,B)$ denotes a normal density with a mean vector A and covariance matrix B]. The vector $\hat{X}(k-1)$ is the mean value of $X(k-1)$ given $Y(k-1),...,Y(1)$. Then the mean value of $X(k)$ given $\hat{X}(k-1)$ is $\Phi\hat{X}(k-1)$. Let the covariance matrix of $X(k-1)$ about $\hat{X}(k-1)$ given all data up to $Y(k-1)$ be $K_{\hat{X}(k-1)}$. (The subscripts on the covariance matrices denote the random variable.) The covariance matrix of $\Phi X(k-1)$ about $\Phi\hat{X}(k-1)$ is $\Phi K_{\hat{X}(k-1)}\Phi^t$ (Φ^t is the transpose of Φ). Since $X(k-1)$ is independent of $U(k)$, we have

$$P[X(k)|\hat{X}(k-1)] = N[\Phi\hat{X}(k-1), \; K_{U(k-1)}\Phi^t + \Phi K_{\hat{X}(k-1)}\Phi^t] \quad (3.4)$$

Assume that $K_{W(k)}$ is not singular. Thus, the product can be written as

$$P[Y(k)|X(k)]P[\dot{X}(k)|\hat{X}(k-1)] = (\text{const}) \exp\left[-\frac{1}{2}\left\{[Y(k)-HX(k)]^t K_{W(k)}^{-1}[Y(k)-HX(k)]\right.\right.$$

$$\left.\left. + [X(k)-\Phi\hat{X}(k-1)]^t K_{X(k)}^{-1}[X(k)-\Phi\hat{X}(k-1)]\right\}\right]$$

(3.5)

where

$$K_{X(k)} \triangleq \Phi K_{\hat{X}(k-1)}\Phi^t + \Gamma K_{U(k-1)}\Gamma^t$$

(3.6)

The exponent will be minimized when the gradient of the quadratic in Eq. (3.5) is zero.

If Q is any symmetric matrix, η is a vector, and if

$$\phi = (HX - \eta)^t Q(HX - \eta)$$

then the gradient of ϕ with respect to X is

$$\text{grad } \phi_X = 2H^t Q(HX - \eta)$$

(3.7)

Also, the gradient of a sum of quadratics is the vector sum of the gradients of each quadratic. If the mode of Eq. (3.5) is found by setting the gradient with respect to $X(k)$ of the exponent equal to zero, and solving for $X(k)$, the optimum estimate of $X(k)$ is:

$$\hat{X}(k) = \left[H^t K_{W(k)}^{-1}H + K_{X(k)}^{-1}\right]^{-1}\left[H^t K_{W(k)}^{-1}Y(k) + K_{X(k)}^{-1}\Phi\hat{X}(k-1)\right]$$

(3.8)

The covariance matrix $K_{\hat{X}(k)}$ is found from Eq. (3.5) by taking the inverse of the sum of the terms that are premultiplied by X^t and postmultiplied by X or

$$K_{\hat{X}(k)} = \left[H^t K_{W(k)}^{-1}H + K_{X(k)}^{-1}\right]^{-1}$$

(3.9)

Note that $K_{\hat{X}(k)}$ is not a function of Y.

If the estimation procedure is stated at $k = 1$, $K_{X(1)}$ must be known. Assume that the dynamical system is turned on at $k = -M_o$ and the initial conditions are zero at that time.

$$X(1) = \sum_{i=0}^{M_o+1} \Phi^{M_o-i+1} \Gamma U(i-M_o-1) \tag{3.10}$$

Then,

$$K_{X(1)} = \sum_{i=0}^{M_o+1} \Phi^{M_o-i+1} \Gamma K_{U(i-M_o-1)} \Gamma^t \left(\Phi^{M_o-i+1}\right)^t \tag{3.11}$$

The one remaining quantity to be specified is $\hat{X}(1)$.

$$P[X(1)|Y(1)] = \frac{P[Y(1)|X(1)]P[X(1)]}{P[Y(1)]} \tag{3.12}$$

$$P[Y(1)|X(1)]\, P[X(1)] = (\text{const}) \exp\left[-\frac{1}{2}\left\{[Y(1) - HX(1)]^t\, K_{W(1)}^{-1}[Y(1) - HX(1)]\right.\right.$$

$$\left.\left. + [X(1) - \overline{X}(1)]^t\, K_{W(1)}^{-1}[X(1) - \overline{X}(1)]\right\}\right] \tag{3.13}$$

where $\overline{X}(1)$ is the a priori mean of $X(1)$ [$\overline{X}(1)$ is zero if the initial conditions in Eq. (3.10) are zero]. Setting the gradient equal to zero gives

$$\hat{X}(1) = \left[H^t\, K_{W(1)}^{-1} H + K_{X(1)}^{-1}\right]^{-1} \left[H^t\, K_{W(1)}^{-1} Y(1) + K_{X(1)}^{-1} \overline{X}(1)\right] \tag{3.14}$$

The matrix $K_{\hat{X}(1)}$ is obtained from Eq. (3.9).

. Equations (3.6), (3.8), (3.9), (3.11), and (3.14) completely specify $\hat{X}(k)$ given $\hat{X}(k-1)$ and $Y(k)$. If it is desired to estimate $X(k)$ given $Y(k),\ldots,Y(1)$, write

$$P[X(k)|Y(k),\ldots,Y(1)] = P[Y(k)|X(k),Y(k-1),\ldots,Y(1)]$$

$$\cdot\ P[X(k)|Y(k-1),\ldots,Y(1)]$$

$$\cdot\ \frac{P[Y(k-1),\ldots,Y(1)]}{P[Y(k),\ldots,Y(1)]} \qquad (3.15)$$

The function $\Phi X(k-1)$ is the mean of $X(k)$ given that $Y(k-1),\ldots,Y(1)$ have occurred. Notice that $K_{X(k)}$ is not a function of $Y(k-1),\ldots,Y(1)$; therefore,

$$P[X(k)|Y(k-1),\ldots,Y(1)] = P[X(k)|\hat{X}(k-1)]$$

Again, it is obvious that

$$P[Y(k)|X(k),Y(k-1),\ldots,Y(1)] \approx N[HX(k),K_{W(k)}] \qquad (3.16)$$

so that Eqs. (3.6), (3.8), (3.9), (3.11), and (3.12) also specify $\hat{X}(k)$ given $Y(k),\ldots,Y(1)$.

In the smoothing problem, it is desired to estimate the value of $X(1)$ given $Y(k),\ldots,Y(1)$. Following a similar method, the conditional density is considered. To obtain the estimate in recursive form, it is desired to express the gradient of $\ln P[X(1)|Y(1),\ldots,Y(k)]$ with respect to $X(1)$ as a function of the gradient of $\ln P[X(1)|Y(1),\ldots,Y(k-1)]$.[*]

$$P[X(1)|Y(1),\ldots,Y(k-1)]$$

$$= \frac{P[X(1),Y(1),\ldots,Y(k-1)]}{P[Y(1),\ldots,Y(k-1)]}$$

$$= \frac{P[X(1)]\ P[Y(1)|X(1)]\ P[Y(2)|X(1),Y(1)],\ldots,P[Y(k-1)|X(1),Y(1),\ldots,Y(k-2)]}{P[Y(1),\ldots,Y(k-1)]}$$

$$(3.17)$$

---

[*]Throughout the balance of this report all gradients will be taken with respect to X.

Since on the right side of Eq. (3.17) $X(1)$ is contained only in the numerator,

$$\text{grad}\left\{\ell n\ P[X(1)|Y(1),\ldots,Y(k-1)]\right\} \propto \text{grad}\left\{\ell n\ P[X(1),Y(1),\ldots,Y(k-1)]\right\}$$

(3.18)

Similarly,

$$\text{grad}\left\{\ell n\ P[X(1)|Y(1),\ldots,Y(k)]\right\} \propto \text{grad}\left\{\ell n\ P[X(1),Y(1),\ldots,Y(k)]\right\}$$

$$= \text{grad}\left\{\ell n\ P[X(1)|Y(1),\ldots,Y(k-1)]\right\}$$

$$+ \text{grad}\left\{\ell n\ P[Y(k)|X(1),Y(1),\ldots,Y(k-1)]\right\}$$

(3.19)

So the gradient conditioned on data to time k is obtained by a simple addition of

$$\text{grad}\left\{\ell n\ P[Y(k)|X(1),Y(1),\ldots,Y(k-1)]\right\}$$

to

$$\text{grad}\left\{\ell n\ P[X(1)|Y(1),\ldots,Y(k-1)]\right\}\ .$$

Setting this new gradient equal to zero and solving for $X(1)$ gives the value of $X(1)$ that maximizes $P[X(1)|Y(1),\ldots,Y(k)]$.

The smoothing equations are given in Table 1. Because of the large amount of algebraic manipulation required, the details of the derivation are reserved for Appendix A. The flow diagram for smoothing is shown in Fig. 3.

C. OBSERVATIONS THROUGH A SECOND DYNAMICAL SYSTEM

Very often the observations will be made through a second dynamical system. For example, if state-variable estimation is used to estimate orbital parameters of a space vehicle, the output of the radio link will be fed to an analog-to-digital converter. These converters have a limited dynamical (amplitude) range, and a narrowband filter usually must be placed after the broadband i-f amplifier to reduce the noise variations. The filter bandwidth may be small enough to influence the statistics of $Y(k)$ (i.e., the noise may no longer be "white" and the covariance matrix of $HX(k)$ will be changed).

TABLE 1. SUMMARY OF THE ESTIMATION EQUATIONS

Filtering	Eq. No.
$\hat{X}(k) = \left[H^t K_{W(k)}^{-1} H + K_{X(k)}^{-1}\right]^{-1} \left[H^t K_{W(k)}^{-1} Y(k) + K_{X(k)}^{-1} \Phi \hat{X}(k-1)\right]$	(3.8)
$K_{\hat{X}(k)} = \left[H^t K_{W(k)}^{-1} H + K_{X(k)}^{-1}\right]^{-1}$	(3.9)
$K_{X(k)} \triangleq \Phi K_{\hat{X}(k-1)} \Phi^t + \Gamma K_{U(k-1)} \Gamma^t$	(3.6)
$K_{X(1)} = \sum_{i=0}^{M_o+1} \Phi^{M_o-i+1} \Gamma K_{U(i-M_o-1)} \Gamma^t \left(\Phi^{M_o-i+1}\right)^t$	(3.11)
$\hat{X}(1) = \left[H^t K_{W(1)}^{-1} H + K_{X(1)}^{-1}\right]^{-1} \left[H^t K_{W(1)}^{-1} Y(1) + K_{X(1)}^{-1} \bar{X}(1)\right]$	(3.14)

Smoothing X(1)	
$\hat{X}(1) = \left[C_{k-1}^t \Phi^t H^t K_{Y(k)}(k\|k-1) H \Phi C_{k-1} + \mathcal{G}_{k-1}\right]^{-1}$ $\cdot \left[C_{k-1}^t \Phi^t H^t K_{Y(k)}(k\|k-1)\left\{Y(k) - H\Phi\hat{X}(k-1)\big|_{X(1)=0}\right\} + J_{k-1}\right]$	(A.8)
$C_{k-1} = \prod_{i=2}^{k-1} \left\{\left[H^t K_{W(i)}^{-1} H + K_{X(i)}^{-1}\right]^{-1} K_{X(i)}^{-1} \Phi\right\}$	(A.5)
$\mathcal{G}_k = \left[C_{k-1}^t \Phi^t H^t K_{Y(k)}^{-1}(k\|k-1) H\Phi C_{k-1} + \mathcal{G}_{k-1}\right]$	(A.11)
$J_k = \left[C_{k-1}^t \Phi^t H^t K_{Y(k)}^{-1}(k\|k-1)\left\{Y(k) - H\Phi\hat{X}(k-1)\big|_{X(1)=0}\right\} + J_{k-1}\right]$	(A.12)
$\mathcal{G}_2 = K_{X(1)}^{-1} + H^t K_{W(1)}^{-1} H + \Phi^t H^t \left[K_{W(2)} + H\Gamma K_{U(2)} \Gamma^t H^t\right]^{-1} H\Phi$	(A.9)
$K_{Y(k)}(k\|k-1) = H\left[\Phi K_{\hat{X}(k-1)} \Phi^t + \Gamma K_{U(k-1)} \Gamma^t\right] H^t + K_{W(k)}$	(A.6)

With No Earlier Data	
$J_2 = \Phi^t H^t \left[K_{W(2)} + H\Gamma K_{U(2)} \Gamma^t H^t\right]^{-1} Y(2) + H^t K_{W(1)}^{-1} Y(1) + K_{X(1)}^{-1} \bar{X}(1)$	(A.10)

With Earlier Data	
$J_2 = \Phi^t H^t \left[K_{W(2)} + H\Gamma K_{U(2)} \Gamma^t H^t\right]^{-1} Y(2) + H^t K_{W(1)}^{-1} + K_{X(1)}^{-1} \Phi\hat{X}(0)$	
where $\hat{X}(0)$ is the estimate of X(0) given all the data to k = 0.	

FIG. 3.   FLOW DIAGRAM FOR SMOOTHING.
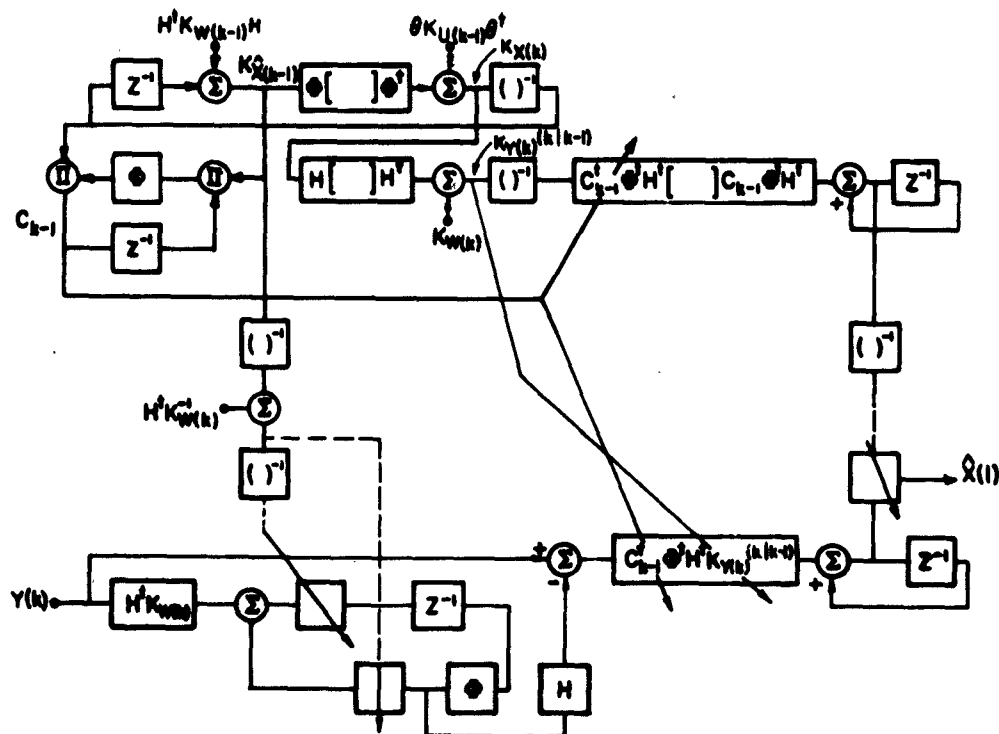
In this section, methods for finding the optimum estimates of $X(k)$ of a function of the second dynamical system will be given. The two dynamical systems will be combined and a single set of system equations will be written. Noise will be added to the output of the second system, and it will be shown that, as the variance of this noise is reduced to zero, the variance of $\hat{X}(k)$ approaches the minimum mean-squared error of $\hat{X}(k)$ in a known manner. Then, some small output noise may be assumed and an estimator may be designed using the combined system equations and the estimation equations of the last section. This estimator may have a mean-squared error as close to the minimum as desired.

Consider the two dynamical systems shown in Fig. 4. The matrix Δ is an input matrix, G is a transition matrix, and J is an output matrix (or vector). Define the following quantities:

$$M(k) = \begin{bmatrix} X(k) \\ X(k) \end{bmatrix} \qquad (3.20)$$

$$\Theta = \begin{bmatrix} G & \vdots & \Delta H \\ \hdots & \vdots & \hdots \\ 0 & \vdots & \Phi \end{bmatrix} \qquad (3.21)$$

$$\gamma = \begin{bmatrix} 0 & \vdots & \Gamma \\ \hdots & \vdots & \hdots \\ \Delta & \vdots & 0 \end{bmatrix} \qquad (3.22)$$

$$N(k) = \begin{bmatrix} W(k) \\ \\ U(k) \end{bmatrix} \qquad (3.23)$$

$$B' = \begin{bmatrix} J & \vdots & 0 \\ \hdots & \vdots & \hdots \\ 0 & \vdots & 0 \end{bmatrix} \qquad (3.24)$$

Then, the system equations for the double dynamical system are

$$M(k) = \Theta M(k-1) + \gamma N(k-1) \qquad (3.25)$$

$$Y(k) = B'M(k) + \Omega(k) \qquad (3.26)$$

The vector $\Omega(k)$ is independent of $\Omega(j)$ for $k \neq j$, and the elements of $\Omega(k)$ are assumed gaussian. The elements of $K_{\Omega(k)}$, a diagonal matrix, are assumed to have a known, very small, upper bound. This small noise generator, of course, is always present in physical situations but its parameters are usually unknown. If, for design purposes, we choose a $K_{\Omega(k)}$ arbitrarily whose elements are the bounds, we are assured by the following argument that the mean-squared error in estimating $M(k)$ from Y is not greater than that error calculated assuming the noise covariance was at the bound. Thus, a bound on performance can be found from the bound on noise power.

The block diagram for the error

$$E(k) = [M(k) - \hat{M}(k)] \qquad (3.27)$$

is shown in Fig. 5. Let $\hat{M}(k)$ be the output of the best linear estimator for $K_{\Omega(k)}$ equal to its upper bound. The error covariance matrix is a sum of a linear (matrix) transformation on $K_{\Omega(k)}$ and a linear
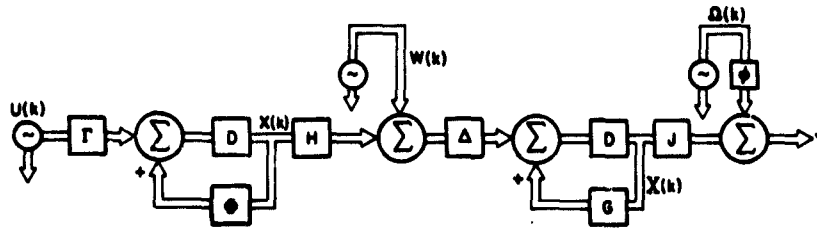
FIG. 4. THE MODEL CONTAINING TWO DYNAMICAL SYSTEMS.
        D is a unit delay.
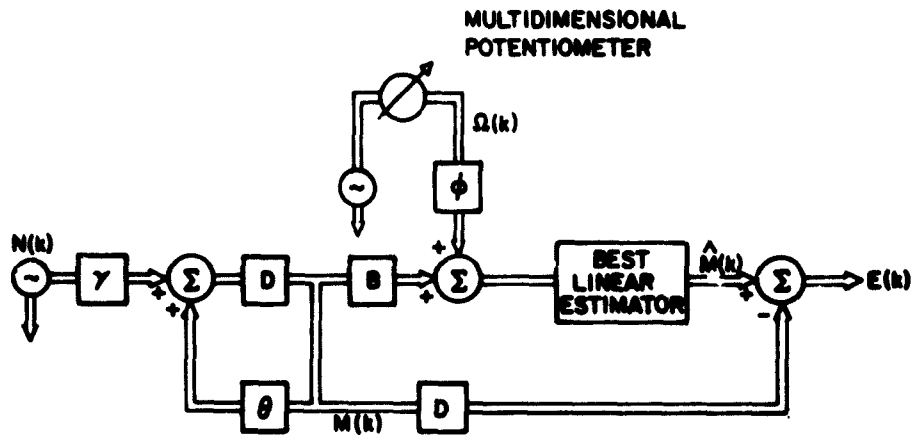
FIG. 5. THE DYNAMICAL SYSTEM FOR FINDING THE ERROR, $E(k)$.

transformation on $K_{N(k-1)}$. The mean-squared error is the trace of this
transformation which is the sum of the trace of the transformation on
$K_{\Omega(k)}$ and the trace of the transformation on $K_{N(k-1)}$. Reduction in
the elements of $K_{\Omega(k)}$ will decrease the mean-squared error since each
diagonal element of the transformation on $K_{\Omega(k)}$ is greater than or
equal to zero. Then, the designer knows that he can lower the upper
bound until the mean-squared error is within specifications.

If the physical situation assures him that the bound is higher than
is actually the case, the estimation equations of Sec. B may be used
directly by substitution of $N(k)$, γ, Θ, B', $\phi\Omega(k)$, and $M(k)$ for
$U(k)$, Γ, F, H, $W(k)$, and $X(k)$, respectively.

D. ESTIMATION WITH PARTIAL DATA

In this section is given the form of the optimum linear estimate when some of the data are missing. Many practical situations arise in which a complete set of observations is not available. An example would be estimation of a rocket trajectory where telemetry was temporarily lost. Another example would be when telemetry is lost during a midcourse maneuver. Frequently the system output is telemetered on a time-shared basis, so that the data are available only periodically.

Assume that all the data are available up to time k and are lost at time $(k+1)$. Then $\hat{X}(k)$ may be calculated using Eq. (3.8). Note that

$$X(k+i) = \phi^i X(k) + \sum_{r=0}^{i-1} \phi^r \Gamma U(j+r) \tag{3.28}$$

Taking the expected values conditioned on $Y(1),\ldots,Y(k)$ [since the $U(j+r)$ have zero mean and are independent of $Y(1),\ldots,Y(k)$] gives

$$\hat{X}(k+i) = \phi^i \hat{X}(k) \tag{3.29}$$

The covariance of $\hat{X}(k+i)$, $K_{\hat{X}(k+i)}(k+i|k)$, conditioned on $Y(1),\ldots,Y(k)$, is

$$K_{\hat{X}(k+i)}(k+i|k) = \phi^i K_{\hat{X}(k)}(\phi^i)^t + \sum_{r=0}^{i-1} \phi^r \Gamma K_{U(j+r)}(\phi^r \Gamma)^t \tag{3.30}$$

Assume that the data are regained at time $(k+j)$. Then the best estimate of $X(k+i)$ in the interval from $(k+1)$ to $(k+j-1)$ is given by Eq. (3.29).

Since new data are available at time $(k+j)$, write

$$P[X(k+j)|Y(k+j),Y(k),\ldots,Y(1)]$$

$$= \frac{P[Y(k+j)|X(k+j)]P[X(k+j)|Y(k),\ldots,Y(1)]P[Y(k),\ldots,Y(1)]}{P[Y(k+j),Y(k),\ldots,Y(1)]} \tag{3.31}$$

and

$$P[X(k+j)|Y(k),\ldots,Y(1)] = N[\phi^j \hat{X}(k); K_{\hat{X}(k+j)}(k+j|k)] \tag{3.32}$$

- 23 -   SEL-64-131

Then the optimum estimate of $X(k+j)$ is

$$\hat{X}(k+j) = \left[HK_{W(k+j)}^{-1}H + K_{\hat{X}(k+j)}^{-1}(k+j|k)\right]^{-1}$$

$$\cdot \left[H^t K_{W(k+j)}^{-1} Y(k+j) + K_{\hat{X}(k+j)}^{-1}(k+j|k)\Phi^j\hat{X}(k)\right] \quad (3.33)$$

The covariance of $\hat{X}(k+j)$ conditioned on the new data is now

$$K_{\hat{X}(k+j)} = \left[H^t K_{W(k+j)}^{-1}H + K_{\hat{X}(k+j)}^{-1}(k+j|k)\right]^{-1} \quad (3.34)$$

The estimation is continued using Eqs. (3.6), (3.8), and (3.9) until a new loss in data occurs. Then Eqs. (3.30), (3.33), and (3.34) are used again.

If smoothing is to be performed, Eq. (A.13) will, of course, not contain the missing data; i.e.,

$$P[X(k+j)|X(1),Y(2),\ldots,Y(k)] = N\left[H\Phi^j\hat{X}(k); HK_{\hat{X}(k+j)}(k+j|k)H^t + K_{W(k+j)}\right] \quad (3.35)$$

The new gradient added by $Y(k\pm j)$ is

$$-2C_k^t(\Phi^j)^t H^t\left[HK_{\hat{X}(k+j)}(k+j|k)H^t + K_{W(k+j)}\right]^{-1}[Y(k+j) - H\Phi^j\hat{X}(k)]$$

Then

$$\hat{X}(1) = \left\{C_k^t(\Phi^j)^t H^t\left[HK_{\hat{X}(k+j)}(k+j|k)H^t + K_{W(k+j)}\right]^{-1}H\Phi^j C_k + \mathcal{G}_k\right\}^{-1}$$

$$\cdot \left\{C_k^t(\Phi^j)^t H\left[HK_{\hat{X}(k+j)}(k+j|k)H^t + K_{W(k+j)}\right]^{-1}[Y(k) - H\Phi^j\hat{X}(k)] + J_k\right\} \quad (3.36)$$

When the next observation, $Y(k+j+1)$ arrives, Eq. (A.8) is used. It should be noted that

$$C_{k+j} = \Phi^j C_k \quad (3.37)$$

and that

$$C_{k+j+1} = \left\{\left[H^t K_W^{-1} + K_{X(k+j+1)}^{-1}\right]^{-1}K_{X(k+j+1)}^{-1}\Phi\right\}\Phi^j C_k \quad (3.38)$$

E. AN EXAMPLE OF A SMOOTHING ESTIMATION

A simple (but typical) dynamical system will be assumed. The matrices will all be scalar.

$$\Phi = \frac{1}{2}, \qquad H = 1, \qquad \Gamma = 1$$

The covariances will be

$$K_{U(k)} = 1, \qquad K_{W(k)} = \frac{1}{4}$$

It is desired to estimate $X(1)$ based on the observations $Y(1)$, $Y(2)$, and $Y(3)$.

The random numbers $U(k)$ and $W(k)$ were obtained from a gaussian random number table with the variance adjusted to 1 and $1/4$ respectively. The numbers chosen were:

k	U(k)	W(k)
1	0.77	0.41
2	-0.33	-0.11
3	--	-0.06

From Eq. (2.3a),

$$X(k) = \frac{1}{2} X(k-1) + U(k-1)$$

$$Y(k) = X(k) + W(k)$$

Then $X(k)$ and $Y(k)$ are:

k	X(k)	Y(k)
1	1	1.41
2	1.27	1.61
3	0.31	0.25

From Eq. (3.6) and Eq. (3.9),

$$K_{\hat{X}(k)} = \left[4 + K_{X(k)}^{-1} \right]^{-1}$$

$$K_{X(k)} = \frac{1}{4} K_{\hat{X}(k-1)} + 1$$

The variances used in calculating $\hat{X}(k)$ given $Y(1),\ldots,Y(k)$ and $X(1)$ equal to zero are then:

k	$K_{X(k)}$	$K_{\hat{X}(k)}$	
1	0	---	
2	1	0.2	
3	1.05	0.202	
4	1.05	0.202	(steady state)

From (3.8),

$$\hat{X}(k) = K_{\hat{X}(k)} \left[4Y(k) + \frac{1}{2} K_{X(k)}^{-1} \right] \hat{X}(k-1)$$

and

$$\hat{X}(2) \Big|_{X(1)=0} = 0.938$$

$$\hat{X}(3) \Big|_{X(1)=0} = 0.297$$

From Eq. (A.5),

$$C_{k-1} = \prod_{i=2}^{k-1} \left\{ \left[4 + K_{X(i)}^{-1} \right]^{-1} K_{X(i)}^{-1} \cdot \frac{1}{2} \right\}$$

$$C_2 = 0.106$$

From Eq. (A.9),

$$\varsigma_2 = K_{X(1)}^{-1} + 4 + \frac{1}{4} \left[\frac{1}{4} + 1 \right]^{-1}$$

The matrix $K_{X(1)}^{-1}$ in Eq. (A.10) is the covariance matrix of $X(1)$ given that $U(k)$ has been applied long enough for the dynamical system to reach steady state at $k = 1$. From Eq. (2.3a),

$$X(1) = \sum_{i=0}^{\infty} \Phi^i U(-i)$$

Then the steady-state variance of $X(1)$ is

$$K_{X(1)} = \sum_{i=0}^{\infty} \left[\left(\frac{1}{2}\right)^2\right]^i \times 1 = \frac{4}{3}$$

and

$$\mathcal{G}_2 = 5.53$$

From Eq. (A.10)

$$J_2 = \frac{1}{2}\left[\frac{1}{4} + 1\right]^{-1} Y(2) + 4 \times Y(1)$$

$$= 6.12$$

where $\bar{X}(1) = 0$. Then

$$\hat{X}[1|Y(1),Y(2)] = \frac{6.12}{5.53} = 1.11$$

From Eq. (A.6)

$$K_{Y(3)} (3|2) = 1.30$$

From Eq. (A.11),

$$\mathcal{G}_3 \cong 5.532$$

From Eq. (A.12), $J_3 \cong 6.11$. Then

$$\hat{X}[1|Y(1),Y(2),Y(3)] \cong 1.10$$

Equation (3.14) may be used to find $\hat{X}[1|Y(1)] = 1.18$.

The results are summarized in Fig. 6. The estimator reaches steady state after only two observations. Further examination will show that the weights of all Y(k) after Y(2) are very nearly zero. Such very short estimator impulse responses occur in a large number of practical problems with typical dynamical systems. As will be shown in Sec. D of Chapter VI, the estimator may always be truncated after a small number of terms except in the rare case of a dynamical system with an extremely high Q. Usually, short impulse estimators for even these cases may be found by choosing an equivalent model of the dynamical system with a much lower sampling rate.
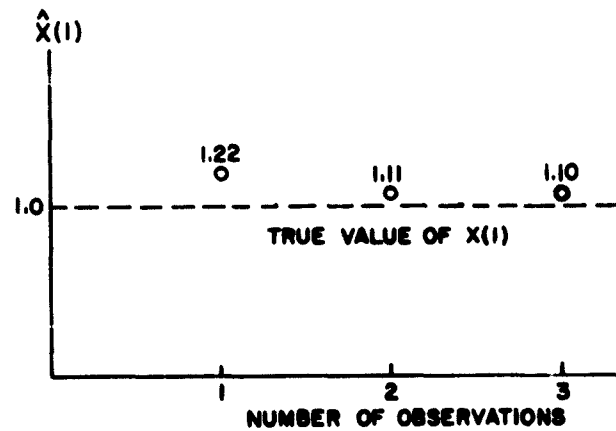
$\hat{X}(1)$

1.22
○

1.11
○

1.10
○

1.0

TRUE VALUE OF X(1)

1   2   3

NUMBER OF OBSERVATIONS

FIG. 6.   THE ESTIMATES OF X(1).

IV. APPLICATIONS TO DETECTION OF GAUSSIAN SIGNALS
IN ADDITIVE GAUSSIAN NOISE

In this chapter the theory of linear state-variable estimation is
applied to the problem of detecting a gaussian signal immersed in additive
gaussian noise. The optimum detector contains an operator that gives
the best estimate of each signal sample value during the detection
interval based on all the data observed during the interval. Typically,
several thousand data points may be observed. In the usual derivation
of the optimum detector (see Sec. A), a matrix of an order equal to the
number of data points must be inverted. When the signal can be repre-
sented as, or approximated by, the output of a noisy dynamical system,
the estimation equations of Chapter III may be applied directly. The
matrices to be inverted will be no larger than the order of the signal-
generating dynamical system, regardless of the number of data points.

Further simplification results if the impulse response of the
estimator of a sample value is truncated when the error covariance matrix
shows that there will be little reduction in mean-squared error by
conditioning on additional data points. A near-optimum time-invariant
detector is then shown to exist.

The chapter begins with a definition of the likelihood ratio. Then
follows a derivation of the optimum detector that requires an inversion
of a matrix of high order. The final section derives the near-optimum
time-invariant detector.

Let the observations $Y(k)$ be

$$Y(k) = x_1(k) + W(k) \qquad\qquad (4.1)$$

when the signal $x_1(k)$ is present (hypothesis ω_1 is true), where $W(k)$
is additive gaussian noise. When no signal is present (hypothesis ω_2
is true),

$$Y(k) = W(k) \qquad\qquad (4.2)$$

The quantities $Y(k)$, $x_1(k)$, and $W(k)$ are assumed to be scalar.

When a statistical decision is made between the presence of a signal in noise, or noise alone, the best decision is based on the likelihood-ratio test [Ref. 10, p. 318]. This ratio is defined as

$$L[Y(1),\ldots,Y(n)] = \frac{P[Y(1),\ldots,Y(n)|\omega_1]}{P[Y(1),\ldots,Y(n)|\omega_2]} \qquad (4.3)$$

If

$$L[Y(1),\ldots,Y(n)] > \beta \qquad (4.4a)$$

we say a signal is present, and if

$$L[Y(1),\ldots,Y(n)] < \beta \qquad (4.4b)$$

we say there is noise alone.

Assume the signal is gaussian with "zero mean" covariance matrix

$$K_{X_n} = E\left\{X_n X_n^t\right\}$$

where

$$X_n = \begin{bmatrix} x_1(1) \\ x_1(2) \\ \vdots \\ x_1(n) \end{bmatrix}$$

Assume that the noise has zero mean and denote the covariance of the signal-plus-noise vector by K_{Y_n}. Then

$$K_{Y_n} = K_{W_n} + K_{X_n} \qquad (4.5)$$

$$P[Y(1),\ldots,Y(n)|\omega_1] = (\text{const})\ \exp\left\{-\frac{1}{2}\ Y_n^t\ K_{Y_n}^{-1}\ Y_n\right\} \qquad (4.6)$$

and

$$P[Y(1),\ldots,Y(n)|\omega_2] = (\text{const})\ \exp\left\{-\frac{1}{2}\ Y_n^t\ K_{W_n}^{-1}\ Y_n\right\} \qquad (4.7)$$

Since the logarithm is a single-valued function, one might just as well consider

$$\ln L[\underline{Y}(1),\ldots,Y(n)] = (\text{const}) - \left\{ Y_n^t \, K_{Y_n}^{-1} \, Y_n - Y_n^t \, K_{W_n}^{-1} \, Y_n \right\} \quad (4.8)$$

Or, a signal is said to be present if

$$- Y_n^t \, K_{Y_n}^{-1} \, Y_n + Y_n^t \, K_{W_n}^{-1} \, Y_n > \gamma_o \quad\quad (4.9)$$

It is noticed that the dimension of Y_n equals the number of data points used in the decision. The inversion of K_{Y_n} may be difficult or impossible in problems involving a great deal of data.

A. THE LIKELIHOOD DETECTOR[*]

This section describes how to calculate the left side of inequality (4.9). It will be shown shortly that this calculation requires finding optimum estimates of the signal conditioned on the observed data. First, the optimum smoother will be derived in a form different from that derived in Chapter III.

A linear estimate of $x_1(i)$ given Y_n will have the form

$$\hat{x}_1(i) = \sum_{j=1}^{n} a_{ij} \, Y(j) \quad\quad (4.10)$$

The error is

$$e(i) = x_1(i) - \sum_{j=1}^{n} a_{ij} \, Y(j) \quad\quad (4.11)$$

and the mean-squared error is

$$\overline{e^2(i)} = R_x(0) - 2 \sum_{j=1}^{n} a_{ij} R_x(i-j) + \sum_{k=1}^{n} \sum_{j=1}^{n} a_{ij} a_{ik} R_{x+w}(j-k) \quad (4.12)$$

---

[*]The following treatment is due to T. Kailath [Ref. 8].

where

$$R_x(i-j) = E[x_1(i)x_1(j)] \qquad (4.13)$$

and

$$R_{x+w}(j-k) = E\left\{[x_1(j) + W(j)][x_1(k) + W(k)]\right\} \qquad (4.14)$$

To minimize $\overline{e^2(i)}$, take the partial derivative

$$\frac{\partial e^2(i)}{\partial a_{ij}} = 0 = 2R_x(i-j) + 2\sum_{k=1}^{n} a_{ik}R_{x+w}(j-k) \qquad j = 1,2,\ldots,n \qquad (4.15)$$

Thus there are n simultaneous equations giving a solution for the n a_{ij}. If this process is repeated for each i, the result may be written as a matrix equation

$$[a_{ij}] \, K_{Y_n} = K_{X_n}$$

which has the solution

$$A = K_{X_n} K_{Y_n}^{-1} \qquad (4.16)$$

The dot product of the ith row of A with Y_n will give the minimum mean-squared-error linear estimate of $x_1(i)$, given Y_n, or

$$AY_n = \begin{bmatrix} \hat{x}_1(1) \\ \vdots \\ \hat{x}_1(n) \end{bmatrix} \triangleq \hat{X}_n \qquad (4.17)$$

Notice that

$$A = \left(K_{Y_n} - K_{W_n}\right)K_{Y_n}^{-1}$$
$$= I - K_{W_n} K_{Y_n}^{-1} \qquad (4.18)$$

SEL-64-131   - 32 -

Then

$$K_{Y_n}^{-1} = K_{W_n}^{-1} - K_{W_n}^{-1} A \qquad (4.19)$$

Substituting Eq. (4.19) into Eq. (4.9), we can now say that a signal is present if

$$Y_n^t K_{W_n}^{-1} A Y_n > \gamma_o \qquad (4.20a)$$

or

$$Y_n^t K_{W_n}^{-1} \hat{X}_n > \gamma_o \qquad (4.20b)$$

As the length of the sample of signal and noise or the size of n grows, so does the dimension. In practice, it is impossible to invert matrices of dimension greater than four or five hundred. In a typical planetary radar detection problem, the signal sample may be 30 min long with a bandwidth of 5 cps. Then, the dimension of K_{Y_n} will be 1.8×10^4. If such problems are to be solved optimally, a more efficient design procedure must be found.

B. A NEAR-OPTIMUM DETECTOR CONTAINING A TIME-INVARIANT FILTER

The remainder of this chapter will show that, for $Y(k)$ stationary and n sufficiently large, there exists another matrix representing a time-invariant filter whose mean-squared error averaged over i is arbitrarily close to the mean-squared error of A averaged over i. Furthermore, the norm of the difference between the ith row of this new matrix and the ith row of A tends toward zero as i increases. As shown in Appendix B, this implies that, as m increases, the detection error probabilities using the time-invariant filter are arbitrarily close to those of the detector employing the filter represented by A.

The linear smoothing equations of Chapter III may be used to find the time-invariant filter. It will be shown that only a small sequence of the elements of Y_n are required at one time, so that computer storage requirements may be considerably reduced.

- 33 -

Assume that a filter with impulse response is represented by the vector $\vec{A}^{(i)}$. Let this filter give the optimum estimate of $x_1(i)$; i.e.,

$$\hat{x}_1(i) = [\vec{A}^{(i)}]^t \, Y_n \qquad (4.21)$$

Let another filter with impulse response vector $\vec{B}^{(i)}$ give another estimate of $x_1(i)$. We now prove that a bound on the vector difference between $\vec{A}^{(i)}$ and $\vec{B}^{(i)}$ may be computed from the difference in mean-squared errors of $\vec{A}^{(i)}$ and $\vec{B}^{(i)}$.

Theorem: If $[\vec{A}^{(i)}]^t \, Y_n$ is the minimum mean-squared error estimate of $x_1(i)$, given Y_n, and $[\vec{B}^{(i)}]^t \, Y_n$ is some other linear estimate of $x_1(i)$ with an increase in mean-squared error of $\overline{\Delta e^2}$ over that of $[\vec{A}^{(i)}]^t \, Y_n$, then

$$\overline{\Delta e^2} \geq [\Delta A^{(i)}]^t \, E\left\{ W_n W_n^t \right\} \Delta A^{(i)}$$

$$= \sigma^2_{w(i)} \, [\Delta A^{(i)}]^t \, \Delta A^{(i)} \qquad (4.22)$$

where $E\left\{ W_n W_n^t \right\} = \sigma^2_{w(i)} I$ and $\Delta A^{(i)}$ is defined as

$$\Delta A^{(i)} = \vec{B}^{(i)} - \vec{A}^{(i)}$$

Proof: Let Δa_{ij} be the jth element of $\Delta A^{(i)}$.

$$\overline{\Delta e^2} = \sum_{j=k}^{n} \sum_{k=1}^{n} (a_{ij} + \Delta a_{ij})(a_{ik} + \Delta a_{ik}) \, R_{x+w} \, (j-k)$$

$$- \sum_{j=1}^{n} \sum_{k=1}^{n} a_{ij} a_{ik} \, R_{x+w} \, (j-k) - 2 \sum_{j=1}^{n} \Delta a_{ij} \, R_x \, (j-i)$$

$$= \sum_{j=1}^{n} \sum_{k=1}^{n} \Delta a_{ij} \Delta a_{ik} \, R_{x+w} \, (j-k) + 2 \sum_{j=1}^{n} \sum_{k=1}^{n} \Delta a_{ij} a_{ik} \, R_{x+w} \, (j-k)$$

$$- 2 \sum_{j=1}^{n} \Delta a_{ij} \, R_x \, (j-i) \qquad (4.23)$$

By Eq. (4.15),

$$\overline{\Delta e^2(i)} = \sum_{j=1}^{n} \sum_{k=1}^{n} \Delta a_{ij} \Delta a_{ik} \, R_{x+w}(j-k)$$

$$= [\Delta A^{(i)}]^t \, K_{Y_n} \, \Delta A^{(i)}$$

$$= [\Delta A^{(i)}]^t \left[K_{X_n} + K_{W_n} \right] \Delta A^{(i)}$$

$$\geq [\Delta A^{(i)}]^t K_{W_n} \, \Delta A^{(i)}$$

$$= \sigma_{w(i)}^2 \, [\Delta A^{(i)}]^t \, \Delta A^{(i)}$$

for white noise since K_{X_n} is positive definite.

Next, a limit for $\overline{e^2(i)}$ given $Y(i-m),\ldots,Y(i+m)$ as $m \to \infty$ will be derived. For $m < \infty$, the mean-squared error will be larger than this limit by some arbitrarily small amount for arbitrarily large m. The length of the impulse response of the filter represented by the ith row of A is n. If the optimum filter with an impulse response 2m long has a mean-squared error $\overline{\Delta e^2}$ greater than the limit, then the theorem above assures that the norm of the difference between the ith row of A and this second filter is bounded by

$$\sqrt{\overline{\Delta e^2}/\sigma_{w(i)}^2}$$

provided $n > 2m$. This bound will be used in Appendix B to show that the difference in error probabilities between the detector of Sec. A (whose estimator has an impulse response of length n) and a detector containing an estimator with an impulse response $2m < n$ long, is also bounded.

Most of the proof of the theorem is used to show that the mean-squared error of the estimate of $x_1(i)$ conditioned on $Y(i-m),\ldots,Y(i+m)$ (found by the methods of Chapter III or Sec. A) is identically equal to the steady-state mean-squared error of the sampled-data smoothing filter

(estimating back for a fixed increment m) with an impulse response
2m long derived using the more classical spectral-analysis approach.
This fact fits one's intuition, but it is not a trivial problem.

Theorem: As $m \to \infty$,

$$\overline{e^2(i)} = R_x(0) - \sum_{j=i-m}^{i+m} a_{ij} R_x(i-j) \to R_x(0) - \int_{-\infty}^{\infty} S_{X+W}(f) \left| \frac{S_X(f)}{S_{X+W}(f)} \right|^2 df$$

$$(4.24)$$

where $S_X(f)$ and $S_{X+W}(f)$ are the signal and noise power spectra,
respectively.

Proof: Consider the equation for the steady-state mean-squared
error of a sampled-data filter with an impulse response 2m long.

$$\overline{e^2} = R_x(0) + \frac{T}{2\pi j} \oint_{\Gamma_o} \left\{ S_{X+W}(Z) \left[\sum_{j=-m}^{m} a_j Z^{-j} \sum_{k=-m}^{m} a_k Z^k \right] \right.$$

$$\left. - S_X(Z) \sum_{j=-m}^{m} a_j (Z^{-j} + Z^j) \right\} \frac{dZ}{Z}$$

$$(4.25)$$

where $S_X(Z)$ and $S_{X+W}(Z)$ are the sampled signal and signal-plus-
noise spectra, and Γ_o is taken to be around the unit circle. Since
$Y(t)$ is bandlimited and if the a_j are picked to minimize the mean-
squared error, $\overline{e^2}$ will approach the mean-squared error of the
optimum continuous noncausal filter as $m \to \infty$. This latter mean-
squared error is well known [Ref. 11] and is

$$R_x(0) - \int_{-\infty}^{\infty} S_{X+W}(f) \left| \frac{S_X(f)}{S_{X+W}(f)} \right|^2 df$$

Equation (4.25) may be rewritten as

$$\overline{e^2} = R_x(0) + \frac{T}{2\pi j} \oint_{\Gamma_o} \left\{ S_{X+W}(Z) \left[\sum_{j=0}^{2m} a_j Z^{-j} \sum_{k=0}^{2m} a_k Z^k \right] \right.$$

$$\left. - S_X(Z) \sum_{j=0}^{2m} a_i (Z^{-j+m} + Z^{j-m}) \right\} \frac{dZ}{Z}$$

$$(4.26)$$

Write

$$\frac{T}{2\pi j} \oint_{\Gamma_o} S_{X+W}(Z) \sum_{j=0}^{2m} a_j Z^{-j} \sum_{k=0}^{2m} a_k Z^k = \frac{T}{2\pi j} \oint_{\Gamma_o} S_{X+W}(Z) \sum_{j=0}^{2m} \sum_{k=0}^{2m} a_j a_k Z^{j-k} \frac{dZ}{Z}$$

$$+ \sum_{j=0}^{2n} a_j \sum_{k=0}^{2n} a_k R_{X+W}[(j-k)T] \qquad (4.27)$$

Using Parseval's theorem for discrete systems gives

$$\frac{T}{2\pi j} \oint_{\Gamma_o} \left[\sum_{j=0}^{2m} a_j (Z^{-j+m} + Z^{j-m}) \right] S_X(Z) \frac{dZ}{Z} = \oint_{\Gamma_o} 2 \left[\sum_{j=0}^{2m} a_j Z^{-j+m} \right] S_X(Z) \frac{dZ}{Z}$$

$$= 2 \sum_{j=0}^{2m} a_j R_X[(j-m)T] \qquad (4.28)$$

In Appendix E, it is shown that $\overline{e^2}$ is minimum if

$$\sum_{k=0}^{2m} a_k R_{X+W}[(j-k)T] = R_X[(jT)]$$

so that

$$\overline{e^2} = R_X(0) - \sum_{j=0}^{2m} a_j R_X[(j-n)T] \qquad (4.29)$$

Putting Eq. (4.15) into Eq. (4.12) gives

$$\overline{e^2(i)} = R_X(0) - \sum_{j=0}^{2m} a_{ij} R_X(j-i) \qquad (4.30)$$

[in which $\overline{e^2(i)}$ is as defined in Eq. (4.12)], or

$$\overline{e^2} = \overline{e^2(i} \qquad (4.31)$$

and the theorem is proved.

If we are able to derive a filter giving an estimate of $x_1(i)$ using as data $Y(i-j),\ldots,Y(i+j),(i-m) > 0$, $(i+m) < n$, with mean-squared error $\overline{e_i^2}(i-m,i+m)$, then we know that the mean-squared error $\overline{e_i^2}(1,n)$, of $\hat{x}_1(i)$, given $Y(1),\ldots,Y(n)$, is bounded by

$$\overline{e_i^2}(i-m,i+m) \geq \overline{e_i^2}(1,n) \geq R_x(0) - \int_{-\infty}^{\infty} S_{X+W}(f) \left| \frac{S_X(f)}{S_{X+W}(f)} \right|^2 df \quad (4.32)$$

And if

$$\overline{e^2} \triangleq \overline{e_i^2}(i-m,i+m) - R_x(0) + \int_{-\infty}^{\infty} S_{X+W}(f) \left| \frac{S_X(f)}{S_{X+W}(f)} \right|^2 df$$

is very small, then $\vec{A}^{(i)}(i-m,i+m)$ will be very close to $\vec{A}^{(i)}(1,n)$, where $\vec{A}^{(i)}(j,\ell)$ is the impulse response of the estimator of $\hat{x}_1(i)$ using data from time equals j to time equals ℓ.

The signal may be a gaussian output of a dynamical system with random inputs, or the signal statistics may always be approximated to any degree by the statistics of such a system. Our signal generator model will be a noisy linear discrete system.

The optimum estimate of a linear operation on a state vector is that same linear operation on the optimum estimate of the state vector. The Bayes estimate of $X(i)$ is the conditional mean of $X(k)$, $\hat{X}(i)$; and the conditional mean of $x_1(i) = HX(i)$ is

$$\hat{x}_1(i) = H\hat{X}(i)$$

Equation (A.8) (smoothing with earlier data) may be used to find $\hat{X}(i)$ and $\hat{x}_1(i)$, given $Y(i-m),\ldots,Y(i+m)$. The error covariance matrix, $K_{\hat{X}(i)}(i-m,i+m)$ may be easily calculated. Then

$$\overline{e_i^2}(i-m,i+m) = HK_{\hat{X}(i)}(i-m,i+m) H^t \quad (4.33)$$

The value of m is increased until the upper bound of Eq. (4.32) approaches sufficiently close to the limit of the second theorem. This will fix the length of the impulse response.

Equation (A.8) may be written in the form

$$\hat{x}(i) = \sum_{j=-m}^{m} E_j Y(i+j) \qquad (4.34)$$

The scalar weight of $Y(i+j)$ in the estimate of $\hat{x}_1(i)$ is

$$a_{ij} = H E_j$$

Notice that E_j is a function of m <u>only</u> and not of i. Thus, the near-optimum filter is <u>time invariant</u>.

The calculation of E_j is greatly simplified by noting that, in Eq. (4.16), the elements of $\hat{A}^{(i)}(i-m,i+m)$ are symmetrical about i. Then, only the first $(m+1)$ E_j need be calculated, and it is much easier to find these than to calculate the last m E_j directly. Examination of Eqs. (A.4), (A.8), (A.11), and (A.12) shows that

$$E_i = G_{2m+1}^{-1} H^t K_W^{-1}$$

$$E_{i-1} = G_{2m+1}^{-1} K_{X(i)}^{-1} \Phi K_{\hat{X}(i)} H^t K_W^{-1}$$

$$E_{i-2} = G_{2m+1}^{-1} K_{X(i)}^{-1} \Phi K_{\hat{X}(i)} K_{X(i)}^{-1} K_{\hat{X}(i-1)} H^t K_W^{-1}$$

$$\vdots$$

$$E_{i-j} = G_{2m+1}^{-1} K_{X(i)}^{-1} \Phi K_{\hat{X}(i)} K_{X(i)}^{-1} K_{\hat{X}(i-1)} \cdots$$

$$K_{\hat{X}(i-j+1)} H^t K_W^{-1} \qquad (4.35)$$

Examination of Eqs. (A.4) and (A.6) shows that

$$K_{\hat{X}(i)}(i-m,i+m) = \left[\Phi^t H^t K_{W(i)}^{-1} \Phi H + K_{X(i)}^{-1} \right.$$

$$\left. + \sum_{j=2}^{m} C_{j-1}^t Q_{j-1}^t \Phi^t H^t K_{Y(j)}^{-1}(j|j-1) H \Phi Q_{j-1} C_{j-1} \right]^{-1} \qquad (4.36)$$

where

$$Q_{j-1} = K_{\hat{X}(j-1)} \, K^{-1}_{X(j-1)} \, \Phi \qquad (4.37)$$

The procedure then is to calculate $\overline{e_1^2}(i-m,i+m)$ for successively larger values of m until $\overline{\Delta e^2}$ is small. The weighting coefficients are calculated according to Eq. (4.35), and the detector is connected as shown in Fig. 7. Notice that the first and last $m \, \hat{x}_1(i)$ are neglected. These, of course, could be calculated; however, for $n \gg m$, little is to be gained by the additional information.
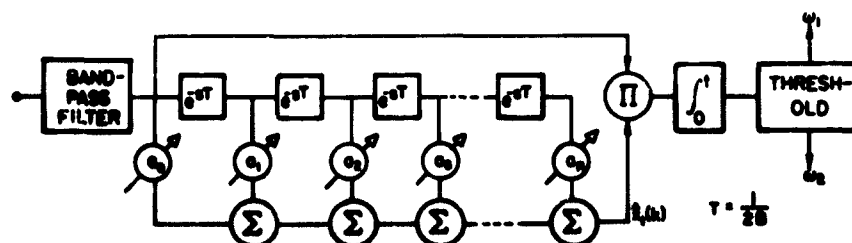

FIG. 7. THE NEAR-OPTIMUM DETECTOR.

V. LINEAR FILTERING OF SIGNALS WITH CONTINUOUS UNKNOWN PARAMETERS

Magill [Ref. 5] has given the form of the general solution to the estimation problem with unknown parameters, and has developed practical methods of implementation for gaussian inputs and a finite parameter set. No general implementation has been developed when the parameters may be chosen from an infinite set or for the optimum linear filter with unknown parameters and nongaussian inputs. This chapter modifies Magill's solution to allow practical estimation of the state variable or signals with a dense set and stationary or near-stationary processes. The distribution over the parameter set is not restricted to gaussian in Magill's solution or in this chapter.

A. MAGILL'S SOLUTION FOR A FINITE NUMBER OF PARAMETER VALUES

If the state of nature ω is to be estimated, the Bayes estimate for a mean-squared loss function is

$$\hat{\omega} = E[\omega|D'] = \int_{\Omega} \omega \, P(\omega|D') \, d\omega \qquad (5.1)$$

where D' is the data. If α is a parameter or parameter vector belonging to the parameter set A_o,

$$P(\omega|D') = \int_{A_o} P(\omega|D',\alpha) \, P(\alpha|D') \, d\alpha \qquad (5.2)$$

and

$$\hat{\omega} = \int_{\Omega} \omega \int_{A_o} P(\omega|D',\alpha) \, P(\alpha|D') \, d\alpha \, d\omega \qquad (5.3)$$

On defining

$$\hat{\omega}(\alpha) \triangleq \int_{\Omega} \omega \, P(\omega|D^r,\alpha) \, d\omega \qquad (5.4)$$

and interchanging the order of integration, Eq. (5.1) is found to be

$$\hat{\omega} = \int_{A_o} \hat{\omega}(\alpha) \; P(\alpha|D') \; d\alpha \qquad (5.5)$$

In other words, the best estimate of ω is the estimate that assumes α is true-weighted by the probability of α conditioned on the data and integrated over A_o.

Magill constructs his adaptive filter by building a number of estimators—one for each member of the parameter set. Then $P[Y(1),\ldots,Y(k)|\alpha_i]$ is evaluated for each α_i. The outputs of the $\hat{\omega}(\alpha_i)$ are weighted by

$$P(\alpha_i|D') = P[\alpha_i|Y(1),\ldots,Y(k)] = \frac{P[Y(1),\ldots,Y(k)|\alpha_i] \; P(\alpha_i)}{P[Y(1),\ldots,Y(k)]}$$
$$(5.6)$$

and summed giving $\hat{\omega}$. Practical methods for evaluating $P[Y(1),\ldots,Y(k)|\alpha_i]$ in the state-variable problem have not been worked out for the dynamical system whose inputs are nongaussian. The best linear $\hat{\omega}(\alpha)$ can still be chosen, however.

B. FILTERING OF STATIONARY OR NEAR-STATIONARY PROCESSES WITH PARAMETERS FROM AN INFINITE SET

The estimator of a state variable or a signal with a given α [the value of the signal or state variable at time k will be denoted by $x_k(k,\alpha)$] is a linear combination of the observed $Y(i)$; i.e.,

$$\hat{x}_1(k,\alpha) = \sum_{j=1}^{k} a_{kj}(\alpha) \; Y(j) \qquad (5.7)$$

From Eq. (5.5)

$$\hat{x}_1(k) = \int_{A_o} \sum_{j=1}^{k} a_{kj}(\alpha) \; P[\alpha|Y(1),\ldots,Y(r)] \; Y(j) \; d\alpha$$
$$= \sum_{j=1}^{k} Y(j) \int_{A_o} a_{kj}(\alpha) \; P[\alpha|Y(1),\ldots,Y(r)] \; d\alpha \qquad (5.8)$$

In other words, the optimum weight is just the mean weight conditioned on the observed data.

In Appendix C a filter is derived that has the form

$$\hat{x}_1(k,\alpha) = \sum_{j=k-k_o}^{k+k_o} a_{kj}(\alpha) \, Y(j) \qquad (5.9)$$

The theory of the last chapter shows that the estimate given by this filter is arbitrarily close to the minimum mean-squared-error estimate for stationary processes. The weight $a_{kj}(\alpha)$ is given by

$$a_{kj}(\alpha) = \sum_{i=1}^{n} a_{kj}^{(i)} \frac{S_i(\alpha)}{S_i(\alpha) + N_i} \qquad (C.33)$$

where S_i and N_i are the signal and noise power in the frequency range $[(B/n)(i-1)]$ to $[(B/n)i]$ cps. The signal is assumed to be in the range from zero to B cps. The constant is found from

$$a_{kj}^{(i)} = a_{ij}' - a_{i(j-1)}' \qquad (C.32)$$

$$a_{ij}' = \frac{\sin\left[\frac{2\pi B}{n} i(jT - k_o T)\right]}{\pi(jT - k_o T)} ; \qquad T = \frac{1}{2B} \qquad (C.31)$$

The noise-power spectrum is assumed to be known so that there is correspondence between the elements of the parameter set and the possible power spectra of $x_1(k,\alpha)$. Then the optimum weight for estimating $x_1(k)$ will be

$$a_{kj} = \sum_{i=1}^{n} a_{kj}^{(i)} \int_{S_i} \frac{S_i}{S_i + N_i} \, P[S_i | Y(1),\ldots,Y(r)] \, dS_i \qquad (5.10)$$

where the conditional density of S_i is the sum of all the densities of $[\alpha : S_i(\alpha) = S_i]$.

Write G_i in transfer matrix form with $q_i(j)$ as the output; i.e.,

$$
\begin{bmatrix} q_i(1) \\ q_i(2) \\ \vdots \\ q_i(r) \end{bmatrix} = G_i \begin{bmatrix} Y(1) \\ Y(2) \\ \vdots \\ Y(r) \end{bmatrix}
\tag{5.11}
$$

The matrix G_i is causal and therefore triangular below the main diagonal. Such matrices are nonsingular if there are no zeros on the main diagonal. Since G_i is time invariant, this is always true, and there will be a one-to-one correspondence between $[Y(1),\ldots,Y(r)]$ and $[q_i(1),\ldots,q_i(r)]$. Thus,

$$
P[S_i | Y(1),\ldots,Y(r)] = P[S_i | q_i(1),\ldots,q_i(r)]
\tag{5.12}
$$

$$
P[S_i | q_i(1),\ldots,q_i(r)] = \frac{P[q_i(1),\ldots,q_i(r)|S_i] \; P[S_i]}{P[q_i(1),\ldots,q_i(r)]}
\tag{5.13}
$$

There are n G_i, each with bandwidth of B/n so that only every n^{th} $q_i(j)$ is needed to find the density of $[q_i(1),\ldots,q_i(r)]$, or

$$
P[S_i | q_i(1),\ldots,q_i(r)] = \frac{P[q_i(1),q_i(n),q_i(2n),\ldots,q_i(r)] \; P[S_i]}{P[q_i(1),q_i(n),q_i(2n),\ldots,q_i(r)]}
\tag{5.14}
$$

The filter G_i is sharp-cutoff so that the $q_i(j)$ in Eq. (5.14) are essentially independent and the following is true:

$$
P[S_i | q_i(1),\ldots,q_i(r)] = \frac{\dfrac{1}{[2\pi(S_i+N_i)]^{r/2n}} \exp\left[-\dfrac{1}{2(S_i+N_i)} \displaystyle\sum_{j=1}^{r/n} q_i^2(nj) \right] P[S_i]}{\displaystyle\int_{S_i} \dfrac{1}{[2\pi(S_i+N_i)]^{r/2n}} \exp\left[-\dfrac{1}{2(S_i+N_i)} \displaystyle\sum_{j=1}^{r/n} q_i^2(nj) \right] P[S_i] dS_i}
$$

$$
\tag{5.15}
$$

The integral in Eq. (5.15) may be evaluated directly as a function of $\sum q_i^2(nj)$ and r, and stored ahead of time.

Since

$$\frac{S_i}{S_i + N_i} = 1 - \frac{N_i}{S_i + N_i}$$

only the following need be calculated:

$$\int_{S_i} \frac{N_i}{S_i + N_i} P[S_i | q_i(1), \ldots, q_i(r)] \, dS_i = \frac{N_i}{(\sqrt{2\pi})^{r/n} P[q_i(1), \ldots, q_i(r)]}$$

$$\cdot \int_{S_i} \frac{N_i}{(S_i + N_i)^{(1+r/n)/2}} \exp\left[-\frac{1}{2(S_i + N_i)} \sum_{j=1}^{r/n} q_i^2(nj)\right] P[S_i] \, dS_i \qquad (5.16)$$

Equation (5.16) is a function only of $[r, \Sigma q_i^2(nj)]$ and may also be calculated ahead of time and stored.

A less complex procedure usually giving almost as good results is maximum likelihood estimation of $(S_i + N_i)$. This estimate is found by solving for S_i in

$$\frac{\partial \{P[S_i | q_i(1), \ldots, q_i(r)]\}}{\partial S_i} = 0 \qquad (5.17)$$

Then

$$-\frac{r}{2n(S_i + N_i)} + \frac{1}{2(S_i + N_i)^2} \sum_{j=1}^{r/n} q_i^2(nj) = 0$$

and

$$\widehat{(S_i + N_i)} = \frac{n}{r} \sum_{j=1}^{r/n} q_i^2(nj) \qquad (5.18)$$

This estimate may be instrumented by an integrate-and-dump circuit or closely approximated by a square-law device followed by a lowpass filter with a bandwidth of $1/r$.

The variance of the estimate of $(S_i + N_i)$ is [Ref. 10, p. 261]

$$\sigma^2_{est} = 2\left(\frac{n}{r}\right)^2 (S_i + N_i)^2 \qquad (5.19)$$

Take, for instance, a 100-tap line with 20 narrowband filters. This adaptive filter can begin its estimating when r equals 100 and, for all practical purposes, it will be converged to Weiner optimum when r equals 200. If the system bandwidth is 100 cps, this convergence will be obtained at the end of 1 sec of operation.

VI. ESTIMATION WITH NONGAUSSIAN INPUTS

A. INTRODUCTION

In this chapter, the theory of Chapter III is extended to include state-variable estimation when either the inputs to the dynamical system or the output noise or both are nongaussian. The same general approach used in Chapter III (finding the mode of the conditional density) is used. Since this density is no longer gaussian, the mode is not necessarily located at the mean. A unimodal density or one with a unique maximum will be assumed. The state variable giving this maximum, of course, is the maximum likelihood estimate. When Bayes estimates are made, it will further be assumed that the density is symmetric about some point.

Most of the discussion in this chapter is concerned with the propagation of nongaussian statistics through a discrete linear dynamical system. At first glance (with the central limit theorem in mind), one might conclude that the output density of a discrete dynamical system with feedback would converge to a gaussian density since the output is the sum of a large number of independent random variables. If this were the case, then linear estimation would be optimum for one or more of the state variables. Unfortunately (as shown in the next section), this is never true for a time-invariant discrete system. In Sec. B, necessary and sufficient conditions for the output density to be of the "same form" as the input density (i.e., the output density is the input density translated and/or with a change in scale) are also derived.

Section C describes the calculation of the joint density of the state variables and the observations. Section D contains a discussion of methods for finding the estimate and it is shown that near optimum estimation may be often obtained with only a short sequence of observations. The last section describes the estimator's asymptotic behavior as the signal-to-noise ratio is increased.

B. PROPAGATION OF FIRST-ORDER STATISTICS

The output of the linear system will be assumed scalar, or the first-order density of only one of the outputs will be of interest.

Then the output may be written as a linear combination of the input
random variables, i.e.,

$$z(k,j_o) = \sum_{\ell=1}^{m_o} \sum_{j=0}^{j_o} a_{\ell j}\, u_\ell(k-j) \qquad (6.1)$$

As before, it is assumed that the $u_\ell(k)$ are independent and have
zero mean with the additional assumption that for a fixed ℓ, the $u_\ell(k)$
are identically distributed for all k. Also assume that for each ℓ
some $a_{\ell j}$ is not equal to zero (i.e., no impulse response from any of
the inputs to the output is equal to zero for all time).

A theorem will now be proved showing that the cdf of the output of
a time-invariant linear filter converges to a gaussian cdf if and only
if the input is gaussian.

Theorem: Let a stable, time-invariant, discrete linear dynamical
system and its input be described by Eq. (6.1) and by the assumptions
given above. Then the cdf of the system output will converge almost
everywhere as $j_o \to \infty$ to a gaussian cdf if and only if the input
is gaussian.

Proof: It is well known that the output of a linear system is
gaussian if the input is gaussian. So only the "only if" proof
will be given here.

It is assumed that, for a fixed ℓ, all $u_\ell(j)$ are identically
distributed. Then

$$\sum_{j=0}^{j_o} \sum_{\ell=1}^{m_o} \sigma^2\Big[a_{\ell j}\, u_\ell(k-j)\Big] = \sum_{\ell=1}^{m_o} \sigma^2\Big[u_\ell(i)\Big] \sum_{j=0}^{j_o} a_{\ell j}^2 \qquad (6.2)$$

If the system is stable,

If the system is stable,[*]

$$\lim_{j_o \to \infty} \sum_{j=0}^{j_o} a_{\ell j}^2 < \infty \qquad \text{for all} \quad \ell$$

Therefore,

$$\lim_{j_o \to \infty} \sum_{\substack{j=0 \\ j \neq i}}^{j_o} \sum_{\ell=1}^{m_o} \sigma^2\left[a_{\ell j} u_\ell(k-j)\right] < \infty \qquad (6.3)$$

Define $z_{\ell i}(k,j_o)$

$$z_{\ell i}(k,j_o) \overset{\triangle}{=} z(k,j_o) - a_{\ell i} u_\ell(k-i) \qquad (6.4)$$

Using Eq. (6.3) and Ref. 12, page 236, it is seen that $z(k,j_o)$ and $z_{\ell i}(k,j_o)$ converge almost surely as $j_o \to \infty$.

The "Composition and Decomposition Theorem" [Ref. 12, p. 271] states that the sum of the two independent random variables with finite means and variances is gaussian if and only if both variables in the sum are gaussian. Therefore, $z(k,j_o)$ and $z(k,\infty)$ are gaussian only if $u_\ell(k-i)$ is gaussian.

At this point it is appropriate to briefly examine a class of zero mean distributions with finite variances that have the interesting

---

[*]The impulse response of a discrete system is bounded. If it is not bounded, a bounded input (i.e., a step function) will give an unbounded output (some of the coefficients in the series form of the output Z-transform will be unbounded). Of course if

$$\sum_{j=0}^{\infty} |a_{\ell j}| < \infty , \qquad \text{then} \qquad \sum_{j=0}^{\infty} a_{\ell j}^2 < \infty .$$

property of being invariant as they are passed through an arbitrary
discrete linear system. Here, "invariant" means only a change in scale
factor.

Definition: A Stable Law [Ref. 12, p. 326]. The cdf, $F(u)$, is
stable if, to every $a_1 > 0$, $a_2 > 0$; and b_1, there correspond
constants $a_3 > 0$ and b_3 such that

$$F(a_1 u_\ell + b_1) * F(a_2 u_\ell = b_2) = F(a_3 u_\ell + b_3) \qquad (6.5)$$

It is clear the invariant distributions belong to the class of stable
laws since it is required that they do not change with an arbitrary
impulse response that is always positive. By a well-known theorem [Ref.
12, p. 327] the log of the characteristic function of a stable law is
given by

$$\log f_{u_\ell}(t) = ita - b|t|^\alpha \left\{ 1 + i\beta \frac{t}{|t|} \omega(t,\alpha) \right\} \qquad (6.6)$$

where

$$\omega(t,\alpha) = \begin{cases} \tan \frac{\pi}{2}\alpha & \text{if} \quad \alpha \neq 1 \\[2mm] \frac{2}{\pi} \ell n \, |t| & \text{if} \quad \alpha = 1 \end{cases} \qquad (6.7)$$

and

$$a \in \text{the real line}$$

$$b \in (0,\infty)$$

$$0 < \alpha \leq 2$$

$$|\beta| \leq 1$$

The variance of $u_\ell(k)$ may be found by taking the second derivative with respect to t of the antilogarithm of Eq. (6.6) and setting $t = 0$. This derivative is less then infinity only when $\alpha = 2$. When $\alpha = 2$, $\omega(t,\alpha) = 0$. Thus, the characteristic function of invariant cdf's must be of the normal form

$$f_{u_\ell}(t) = \exp\left(ita - b|t|^2\right) \tag{6.8}$$

C. FINDING THE JOINT DENSITY OF THE STATE VARIABLE AND THE OBSERVATIONS

Since

$$P[X(1),\ldots,X(k),Y(1),\ldots,Y(k)]$$

$$= P[X(1),\ldots,X(k)|Y(1),\ldots,Y(k)]\ P[Y(1),\ldots,Y(k)]$$

the $X(1),\ldots,X(k)$ that maximize the density on the left side of the above equation also maximize the conditional density on the right. The approach used here will be to first calculate in a convenient form the joint density of the state variables and the observations for nongaussian inputs. In Chapter III this was relatively simple because the joint density of gaussian variables is gaussian, but for nongaussian inputs, finding the joint density is really the heart of the problem. In the next section methods are given for finding the density maximum (the optimum estimate).

The joint density of the state variables alone is

$$P[X(1),X(2),\ldots,X(k)] = P[X(1)]\ P[X(2)|X(1)]$$

$$\cdot\ P[X(3)|X(1),X(2)]\ \ldots\ P[X(k)|X(1),\ldots,X(k)]$$

$$\tag{6.9}$$

The density of $X(1)$ is just the density of $\Gamma U(0)$, or

$$P[X(1)] = P[\Gamma U(0)] \tag{6.10}$$

The density of X(2) conditioned on X(1) is the density of $\Gamma U(1)$
translated by $\Phi X(1)$. Write this density as*

$$P[X(2)|X(1)] = P[\Gamma U(1)|\text{mean} = \Phi X(1)] \tag{6.11}$$

Because of the Markov property of X(i)

$$P[X(3)|X(1)X(2)] = P[X(3)|X(2)] = P[U(2)|\text{mean} = \Phi X(2)] \tag{6.12}$$

and

$$P[X(k)|X(1),\ldots,X(k-1)] = P[\Gamma U(k-1)|\text{mean} = \Phi X(k-1)] \tag{6.13}$$

Since the elements of U(k-1) are independent, it is quite simple
to write down the right side of Eq. (6.13) if Γ is diagonal. Then

$$P[\Gamma U(k-1)|\text{mean} = X(k-1)] = \prod_i P[\gamma_{ii} u_i(k-1)|\text{mean} = \phi_i(k-1)] \tag{6.14}$$

where $\phi_i(k)$ is the ith element of $\Phi X(k)$.

---

*By this notation, we mean that the zero mean density of the random
variable $\Gamma U(1)$ is shifted by an amount equal to $\Phi X(1)$.

If Γ is triangular, there is dependence between the elements of the noise vector added to $\Phi X(k-1)$ and therefore

$$P[\Gamma U(k)|\text{mean} = \Phi X(k-1)] = P[\gamma_{11}u_1(k-1)|\text{mean} = \phi_1(k-1)]$$

$$\cdot \ P[\gamma_{22}u_2(k-1)|\text{mean} = \phi_2(k-1) + \gamma_{21}u_1(k-1)]$$

$$\cdot \ P[\gamma_{33}u_3(k-1)|\text{mean} = \phi_3(k-1) + \gamma_{31}u_1(k-1)$$

$$+ \ \gamma_{32}u_2(k-1)]$$

$$\vdots$$

$$\cdot \ P\left[\gamma_{m_o m_o}u_{m_o}(k-1)|\text{mean} = \phi_{m_o}(k-1) + \sum_{\ell=1}^{m_o-1} \gamma_{m_o \ell}u_\ell(k-1)\right]$$

$$= P[x_1(1)|\phi_1(k)] \ P[x_2(k)|\phi_2(k-1),x_1(k)] \ \ldots$$

$$P\left[x_{m_o}(k)|\phi_{m_o}(k-1),x_1(k),\ldots,x_{m_o-1}(k)\right] \quad (6.15)$$

Similar relationships may be written if Γ can be partitioned into triangular matrices,[*] or if the rows can be reordered to form a triangular matrix or a matrix that has triangular partitions.

---

[*] Here the matrices that are to be partitioned must not have elements of two different triangles in the same row.

It becomes more complex if no such partitions are possible. For example, suppose

$$\Gamma = \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \qquad (6.16)$$

and none of the γ_{ij} equals zero. Then

$$P[X(1)] = \left\{ P[\gamma_{11}u_1(0)] * P[\gamma_{12}u_2(0)] \right\} P[\gamma_{21}u_2(0)$$
$$+ \gamma_{22}u_2(0)|\gamma_{11}u_1(0) + \gamma_{12}u_2(0)] \qquad (6.17)$$

The random variable $\gamma_{21}u_1(0)$ conditioned on $\gamma_{11}u_1(0) + \gamma_{12}u_2(0)$ is not independent of $\gamma_{22}u_1(0)$ conditioned on the same sum. Therefore, convolution cannot be used to find

$$P[\gamma_{21}u_1(0) + \gamma_{22}u_2(0)|\gamma_{11}u_1(0) + \gamma_{12}u_2(0)]$$

Consider the matrix transformation

$$Z(k) = \Xi^{-1}X(k)$$

Then

$$Z(k) = \Xi^{-1}\Phi X(k-1) + \Xi^{-1}\Gamma U(k-1)$$

$$= \Xi^{-1}\Phi \Xi(k-1) + \Xi^{-1}\Gamma U(k-1) \qquad (6.18a)$$

and

$$Y(k) = H \Xi Z(k) + W(k) \qquad (6.18b)$$

are the equations of a system with state variable $Z(k)$ having a response identical to that of the system described by the equations (2.3a) and (2.3c). If Γ is nonsingular, pick

$$\Xi^{-1} = \Gamma^{-1} \qquad (6.19)$$

Then the input matrix will be diagonal.

If a Bayes estimate of $Z(k)$ is made, as noted in Chapter IV, the Bayes estimate of $X(k)$ is

$$\hat{X}(k) = \Xi\hat{Z}(k) \qquad (6.20)$$

Since the Jacobian of a matrix transformation does not depend on the multiplying vector [Ref. 13], the transform of the maximum likelihood estimate of $Z(k)$ will maximize the conditional density of $X(k)$. So Eq. (6.20) is also true for this latter type of estimate.

If Γ is singular, then a nonsingular Ξ^{-1} can be found that, when multiplied with Γ, gives a triangular product. This is best illustrated by an example. Let

$$\Gamma = \begin{bmatrix} \gamma_{11} & \gamma_{12} & 0 \\ \gamma_{21} & \gamma_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

and let

$$\begin{bmatrix} \xi^{11} & \xi^{12} & \xi^{13} \\ \xi^{21} & \xi^{22} & \xi^{23} \\ \xi^{31} & \xi^{32} & \xi^{33} \end{bmatrix} \begin{bmatrix} \gamma_{11} & \gamma_{12} & 0 \\ \gamma_{21} & \gamma_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} b_{11} & 0 & 0 \\ b_{21} & b_{22} & 0 \\ b_{31} & b_{32} & 0 \end{bmatrix} = B_o$$

where ξ^{ij} is the ij^{th} element of Ξ^{-1}.

The first column of B_o can be arbitrary and the last column is obviously all zero. The elements b_{22} and b_{32} can also be arbitrary, but b_{12} must be zero. Then ξ^{11} and ξ^{12} must be picked so that

$$\xi^{11}\gamma_{12} + \xi^{21}\gamma_{21} = 0$$

All the other ξ^{ij} may be chosen to make Ξ^{-1} nonsingular.

The joint density of the state variables will now be used to derive the joint densities of the state variables with the observations. Two

cases will be considered. The first procedure will use convolution of two independent densities, and is not applicable to all forms of dynamical systems. The second is more general but in some cases requires more calculation.

Case I: H (a vector) is different from zero in one dimension only, or H (a matrix) is diagonal.

Any dynamical system with a scalar output may be put in a form where the output is $\mu x_\ell(k)$ where μ is some constant. For the moment we will consider only these systems, or those where H is a diagonal matrix. If the scalar or an element of a vector output is a linear combination of several of the $x_\ell(k)$'s, convolution usually cannot be used to find the probability of the linear combination conditioned on the $x_\ell(j)$, for j less than k. This is true because, in general, the conditioned $x_\ell(k)$ will not be independent.

For convenience, the treatment in Case I will consider scalar outputs only. Extensions to vector outputs will be obvious. Since $P[X(1),\ldots,X(k)]$ is really a joint density, i.e.,

$$P\{[x_1(1),\ldots,x_{m_o}(1)] \ldots [x_1(k),\ldots,x_{m_o}(k)]\}$$

and $W(1)$ is independent of $x_1(1)$, one can write

$$P[Y(1)|x_2(1),\ldots,x_{m_o}(k)] \, P[x_2(1),\ldots,x_{m_o}(k)]$$

$$= P[W(1)] * P[x_1(1)|x_2(1),\ldots,x_{m_o}(k)] \, P[x_2(1),\ldots,x_{m_o}(k)]$$

$$= P[W(1)] * P[x_1(1),x_2(1),\ldots,x_{m_o}(k)] \tag{6.21}$$

By successive convolution, the joint density

$$P[Y(1),x_2(1),\ldots,x_{m_o}(1), \, Y(2),x_2(2),\ldots,x_{m_o}(2), \, \ldots, \, Y(k-1),\ldots,x_{m_o}(k-1)$$

$$x_1(k),\ldots,x_{m_o}(k)]$$

may be found. If desired, all $x_\ell(j)$, j less than k, may be eliminated by taking the marginal distribution giving

$$P[Y(1),Y(2),\ldots,Y(k-1),x_1(k),\ldots,x_{m_o}(k)]$$

Then, note that

$$P[X(k)|Y(1),Y(2),\ldots,Y(k)] = P[Y(k)|X(k)]P[x_1(k),\ldots,x_{m_o}(k),Y(k-1),\ldots,Y(1)]$$

$$= P[W(k)|\text{mean} = HX(k)]$$

$$\cdot\ P[X_1(k),\ldots,x_{m_o}(k),Y(k-1),\ldots,Y(1)] \qquad (6.22)$$

Equation (6.22) must be used because another convolution would eliminate $x_1(k)$. The convolution usually can be written down by inspection.

    <u>Case II</u>: There is no restriction on H.

    The remainder of the derivation here is very simple.

$$P[Y(1),X(1),\ldots,X(k)] = P[Y(1)|X(1)]P[X(1),\ldots,X(k)$$

$$P[Y(1),Y(2),X(1),\ldots,X(k)] = P[Y(2)|X(2)]P[Y(1),X(1),\ldots,X(k)]$$

$$P[Y(1),\ldots,Y(k),X(1),\ldots,X(k)] = P[Y(k)|X(k)]P[Y(1),\ldots,Y(k-1),X(1),\ldots,X(k)]$$

$$= P[X(1),\ldots,X(k)] \prod_{j=1}^{k} P[Y(2)|X(j)] \qquad (6.23)$$

where

$$P[Y(j)|X(j)] = P[W(j)|\text{mean} = HX(j)] \qquad (6.24)$$

Notice that there are more $x_\ell(j)$ in Eq. (6.24) than in Eq. (6.22), and that there will be more integrations if the marginal densities are found. Then Eq. (6.22) should be used if the dynamical system can be described as in Case I and if marginal densities are to be taken.

One other density will be needed in the next section. Assume that the system is of the type considered in Case I but that $P[Y(1),...,Y(j), X(1),...,X(j)]$ has been calculated by the procedures outlined in Case II.

$$P[Y(1),...,Y(k),X(1),...,X(j)] = P[Y(1),...,Y(j),X(1),...,X(j)$$

$$\cdot P[Y(j+1)|X(j)]P[Y(j+2)|X(j),Y(j+1)] \ldots$$

$$\cdot P[Y(k)|X(j),Y(j+1),...,Y(k-1)] \quad (6.25)$$

Note that

$$X(j+1) = \phi^i X(j) + \sum_{r=0}^{i-1} \phi^r \Gamma U(j+r) \qquad (6.26)$$

The following set of equations can be written:

$$Y(j+1) = x_1(j+1) + W(j+1)$$

$$= (\phi_{11},...,\phi_{1m_o})X(j) + (\gamma_{11},...,\gamma_{1m_o})U(j) + W(j+1)$$

$$Y(j+2) = x_1(j+2) + W(j+2)$$

$$= (\phi_{11}^{(2)},...,\phi_{1m_o}^{(2)})X(j) + (\phi_{11},...,\phi_{1m_o})\Gamma U(j)$$

$$+ (\gamma_{11},...,\gamma_{1m_o})\Gamma U(j+1) + W(j+2)$$

$$Y(j+i) = (\phi_{11}^{(i)},...,\phi_{1m_o}^{(i)})X(j) + (\phi_{11}^{(i-1)},...,\phi_{1m_o}^{(i-1)})\Gamma U(j) + \ldots$$

$$+ (\gamma_{11},...,\gamma_{1m_o})U(j+i-1) + W(j+i) \qquad (6.27)$$

where $\phi_{gt}^{(i)}$ is the gt^{th} element of ϕ^i.

Note that

$$P[Y(j+i)|X(j),Y(j+1),\ldots,Y(j+i-1),U(j),\ldots,U(j+i-1)]$$

$$= P[Y(j+i)|X(j),U(j),\ldots,U(j+i-1)]$$

$$= P[W(j+i)|\text{mean} = (\phi_{11}^{(i)},\ldots,\phi_{1m_o}^{(i)})X(j),\ldots,(\gamma_{11},\ldots,\gamma_{1m_o})U(j+i-1)]$$

$$\tag{6.28}$$

So, using all the lines of Eqs. (6.27) and (6.28), we may find

$$P[Y(j+1),\ldots,Y(j+i),U(j),\ldots,U(j+i-1)|X(j)]$$

$$= P[U(j)]P[U(j+1)]\ldots P[U(j+i-1)] \prod_{n=1}^{i} P[Y(j+n)|X(j),U(j),\ldots,U(j+n-1)]$$

$$\tag{6.29}$$

D. FINDING THE ESTIMATE

This section contains a brief description of several methods for calculating the mode of the joint densities found by methods given in the preceding section. The problem is to find the Bayes estimates of $X(1),\ldots,X(k)$ from the observed data $Y(1),\ldots,Y(k)$. First, several observations are made about the convergence of Bayes estimates of linear-dynamic-system state variables. It is obvious that the mean-squared error of $\hat{X}(j)$ will converge to some value as the number of observations increases, and that the mean-squared error must be a nonincreasing function of $(k-j)$, where the estimate is conditioned on $Y(1)$ to $Y(k)$. If the $(k+1)^{\text{th}}$ observation increased the mean square, then the Bayes estimator would ignore $Y(k+1)$. Since the mean-squared error is never negative, it must approach a limit as $(k-j) \to \infty$.

If the estimator is linear and the noise is white and stationary, the estimator will be stable. For example, let the noise be scalar. Then, the component of the mean-squared error of $x_{\ell}(j)$ due to the noise will have the form

$$\sigma_w^2 \sum_{i=1}^{k} a_\ell^2(i)$$

where $a_\ell(i)$ is the value of the estimator input to $x_\ell(j)$ output impulse response at time equal to i. Since the sum is finite (because the Bayes error is finite), the estimator is stable. Similar reasoning shows that a linear Bayes filter and a Bayes predictor are also stable.

In Chapter III there was an example of a linear smoother that obtained near-optimum performance with a small number of observations. In general, nonlinear estimators also obtain near-optimum performance with a truncated sequence of $Y(k)$.

It will now be shown that the desired length of this truncated sequence can be directly related to the rate of decay of an initial condition in the generating dynamical system.

The optimum estimate may be written as

$$\hat{x}(j) = \int_{x(j)} x(j)\ P[x(j)|Y(1),\ldots,Y(k)]\ dx(j)$$

$$= \int_{U(j)} \cdots \int_{U(k-1)} \int_{x(j)} x(j)\ P[x(j)|Y(1),\ldots,Y(k),U(j),\ldots,U(k-1)]$$

$$\cdot\ P[U(j),\ldots,U(k-1)|Y(1),\ldots,Y(k)]\ dx(j)\ dU(j)\ \ldots\ dU(k-1)$$

$$= \int_{U(j)} \cdots \int_{U(k-1)} \hat{x}[x(j)|Y(1),\ldots,Y(k),U(j),\ldots,U(k-1)]\ P[U(j),\ldots,U(k-1)|Y(1),\ldots,Y(k)]\ dU(j)\ \ldots\ dU(k-1)$$

(6.30)

Then, the magnitude of the change in $\hat{x}(j)$, when it is conditioned on one more observation, $Y(k+1)$, is less than

$$\max_{U(j),\ldots,U(k),Y(1),\ldots,Y(k+1)} \left| E[X(j)|Y(1),\ldots,Y(k+1),U(j),\ldots,U(k)] \right.$$

$$\left. - E[X(j)|Y(1),\ldots,Y(k),U(j),\ldots,U(k-1)] \right|$$

The value of $E[X(1)|Y(1),\ldots,Y(k+1),U(j),\ldots,U(k)]$ may be found by solving the set of simultaneous equations

$$\frac{\partial}{\partial x_1(j)} P[X(j),Y(1),\ldots,Y(k+1),U(j),\ldots,U(k)] = 0$$

$$\vdots$$

$$\frac{\partial}{\partial x_{m_0}(j)} P[X(j),Y(1),\ldots,Y(k+1),U(j),\ldots,U(k)] = 0 \qquad (6.31)$$

Consider the ℓth of those equations. From Eq. (6.29),

$$\frac{\partial}{\partial x_\ell(j)} P[X(j),Y(1),\ldots,Y(k+1),U(j),\ldots,U(k)]$$

$$= \prod_{i=j}^{k} P[U(i)]P[X(j),Y(1),\ldots,Y(j)] \prod_{i=1}^{k-j} P[Y(j+i)|X(j)U(j),\ldots,U(j+i-1)]$$

$$\cdot \frac{\partial}{\partial x_\ell(j)} P[Y(k+1)|X(j)U(j),\ldots,U(k)]$$

$$+ P[Y(k+1)|X(j),U(j),\ldots,U(k)] \frac{\partial}{\partial x_\ell(j)} P[X(j),Y(1),\ldots,Y(j)]$$

$$\cdot \prod_{i=1}^{k-j} P[Y(j+i)|X(j),U(j),\ldots,U(j+i)] = 0 \qquad (6.32)$$

From Eq. (6.29), it is seen that a change in $X(j)$ of $\Delta X(j)$ in $P[Y(k)|X(j),U(j),\ldots,U(k-1)]$ changes only the mean of the conditional density by an amount proportional to

$$\left(\phi_{11}^{(k-j)} \cdots \phi_{1m_0}^{(k-j)} \right) \Delta X(j)$$

In any stable dynamical system,

$$\left\| \left(\phi_{11}^{(k-j)} \quad \cdots \quad \phi_{1m_0}^{(k-j)} \right) \right\| \to 0$$

as $k-j \to \infty$. In practice, it is found that this norm goes to zero very quickly. For example, if a and b in the system of Fig. 1 both equal 1/2, for $k-j = 8$,

$$\left(\phi_{11}^{(8)} \phi_{12}^{(8)} \right) \Delta X(j) = \frac{1}{1024} \Delta x_1(j) + \frac{9}{512} \Delta x_2(j) \tag{6.33}$$

Then, the magnitude of the component of the gradient of $P[Y(k)|X(j)$, $U(k),\ldots,U(k-1)]$, with respect to $X(j)$ in the x_1 direction, will be 1/1024 the maximum magnitude of the slope of $P[W(k)]$ in that direction. Clearly, a good engineering approximation is to set the first term in the sum in Eq. (6.32) equal to zero, then the left side of Eq. (6.32) is equal to zero·at exactly the same places

$$\frac{\partial}{\partial x_\ell(j)} P[X(j),Y(1),\ldots,Y(k),U(j),\ldots,U(k-1) = 0$$

There will be no change in the conditional expected value, and a truncated sequence gives very nearly the optimum estimate.

When actually obtaining the estimates, it is usually convenient to use the logarithm of the joint density. From Eqs. (6.23) and (6.25),

$$\ell n\left\{P[Y(1),\ldots,Y(k),X(1),\ldots,X(j)]\right\} = \ell n\left\{P[X(1)]\right\} + \ell n\left\{P[Y(1)|X(1)]\right\} + \ldots$$

$$+ \ell n\left\{P[X(j)|X(j-1)]\right\} + \ell n\left\{P[Y(j)|X(j)]\right\}$$

$$+ \ell n\left\{P[Y(j+1),\ldots,Y(k)|X(j)]\right\} \tag{6.34}$$

Taking the derivative of Eq. (6.34) with respect to $x_\ell(j)$ for all ℓ gives a set of m_0 equations of the form

$$\frac{\partial\, \ell n\left\{P[X(j)|X(j-1)]\right\}}{\partial x_\ell(j)} + \frac{\partial\, \ell n\left\{P[Y(j)|X(j)]\right\}}{\partial x_\ell(j)} + \frac{\partial\, \ell n\left\{P[Y(j+1),\ldots,Y(k)|X(j)]\right\}}{\partial x_\ell(j)} = 0$$
$$\tag{6.35}$$

If the Bayes $\hat{X}(j-1)$, given $Y(1),\ldots,Y(k)$, is known, then $X(j)$ is the only variable in Eq. (6.35); i.e., the $X(j-1)$ that maximizes (or minimizes)

$$P[X(1),\ldots,X(j-1),Y(1),\ldots,Y(k)]$$

also maximizes

$$P[X(1),\ldots,X(j-1),X(j),Y(1),\ldots,Y(k)]$$

This can be seen from the following argument. The Bayes estimate of $X(j-1)$ obviously maximizes $P[X(1),\ldots,X(j-1),Y(1),\ldots,Y(k)]$, and the Bayes estimate of

$$\eta(j) = \begin{bmatrix} x(j-1) \\ x(j) \end{bmatrix}$$

maximizes $P[X(1),\ldots,X(j),Y(1),\ldots,Y(j)]$. The first n elements of $\hat{\eta}(j)$ are the elements of $\hat{X}(j-1)$.

If surface-searching methods are used to solve Eq. (6.35), the surface has dimension m_0. When $\hat{X}(j)$ is found, it is put in the likelihood functions for $X(j+1)$, and the process is repeated until all k of the $X(j)$ are estimated. The first surface search [in the estimation of $\hat{X}(1)$] will use the initial condition, $X(0)$.

If only filtering is desired, the maximum likelihood estimate may be made by using the density described in the last paragraph of Sec. C, or surface searching may be used to solve equations of the form

$$\frac{\partial \ln\{P[X(j)|Y(1),\ldots,Y(j)]\}}{\partial x_\ell(j)} + \frac{\partial \ln\{P[Y(j)|X(j)]\}}{\partial x_\ell(j)} = 0 \qquad (6.36)$$

Notice that $P[X(j)|Y(1),\ldots,Y(j)]$ has the property seen in Chapter III. The observed $Y(i)$ will appear only in the mean, and the other moments do not depend on $Y(i)$. The surface to be searched is again of dimension m_0.

Also, the density necessary for maximum likelihood smoothing with earlier data may be obtained by writing

$$P[X(j),Y(1),...,Y(k)] = P[X(j)|Y(1),...,Y(j-1)] \ P[Y(j)|X(j)]$$

$$\cdot \ P[Y(j+1),...,Y(k)|X(j)] \qquad (6.37)$$

The surface is also m_o-dimensional.

If linear estimation of the $X(j)$ are made, this should be a good first guess for the surface search. In multiple-mode problems with reasonably good signal-to-noise ratios, it would be an aid in starting on the correct mode.

If the dimension of the input to the dynamical system is smaller than the dimension of the state vector, it may be simpler to use Eq. (6.28) directly and estimate the $U(k)$. This would reduce the dimensions of the surface to be searched, and there would be no need for taking marginal distributions. Then, the optimum estimates of the $\hat{x}(i)$ are given by

$$\hat{X}(2) = \Phi X(1) + \hat{U}(1)$$

$$\vdots$$

$$\hat{X}(j) = \Phi X(j-1) + \hat{U}(j-1)$$

where $X(1)$ is the known initial condition.

Another calculation aid is the fact that, with many input densities and many dynamical systems, $P[Y(k)|X(j),Y(j+1),...,Y(k-1)]$ will appear very gaussian if $k-j$ is greater than 4 or 5 and if $W(k)$ is gaussian. This is especially true for moderately high signals (see the next section), or very high gaussian output noise. Then the methods of Chapter III may be used to calculate the conditional density.

Several short comments are offered on finding the mode when the solution is not put in the form of Eq. (6.35). For example, perhaps only a small number of data points are available and the joint density is put into the form of Eq. (6.23). If none of the $x_\ell(j)$ is integrated

out, and a surface search is run over all the $X(j)$, it is often helpful to note that the mode will be at the origin for all $Y(j)$ equal to zero and $P[u_\ell(j)]$ symmetrical about zero [$W(j)$ is assumed gaussian]. Then the $Y(j)$ can be moved out toward their observed values in small steps, thus moving the mode in small steps from a known position. The surface search will be over a much smaller area and the total calculation will be greatly reduced. If there are multiple modes, the correct one will be tracked.

Once Eqs. (6.23), (6.28), or (6.34) have been found, the surface search is a routine problem for the computer programmer. A common procedure is the method of steepest descent (or ascent). The gradient is evaluated at a point and a move is made in the direction of the gradient to a new point. The gradient is again evaluated and the process continued until the gradient magnitude is small. Gradients of either the log of the density or the density itself may be used. Another common procedure is to move along a coordinate axis until there is no increase in the density. This process is repeated in turn on each axis until the mode is reached.

E. THE ASYMPTOTIC BEHAVIOR OF THE ESTIMATORS AS THE SIGNAL-TO-NOISE RATIO IS INCREASED

This section contains a discussion of the asymptotic behavior of estimates for systems with nongaussian inputs as the signal-to-noise ratio is increased. It will be shown that linear estimation is very nearly optimum for high signal-to-noise ratios for one or more states of the dynamical system. In special cases, it can be shown to be nearly optimum for all states of the system. This property of estimators assumes special importance because, in the majority of cases, there probably will be a strong signal and weak noise.

Denote by $X_o(j)$ all elements of $X(j)$ that may be measured directly or calculated exactly knowing $Y(1),\ldots,Y(k)$, $k \geq j$, when the covariance matrix of W is zero. Using the techniques of Sec. C, one can find $P[X_o(j),Y(1),\ldots,Y(k)]$. Let $\hat{x}_\ell(j)$ be the elements of the $X_o(j)$ estimation, $\hat{X}_o(j)$. Now, take a multidimensional Taylor series expansion of $\ln\{P[X_o(j),Y(1),\ldots,Y(k)]\}$ about $\hat{X}_o(j)$.

- 65 -

$$\ln\{P[X_o(j),Y(1),\ldots,Y(k)]\} = \ln\{P[\hat{X}_o(j),Y(1),\ldots,Y(k)]\}$$

$$+ \frac{1}{2} \sum_{g,h} \left. \frac{\partial^2 \ln\{P[X_o(j),Y(1),\ldots,Y(k)]\}}{\partial x_g(j)\partial x_h(j)} \right|_{\substack{x_g(j) = \hat{x}_g(j) \\ \\ x_h(j) = \hat{x}_h(j)}} [x_g(j) - \hat{x}_g(j)][x_h(j) - \hat{x}_h(j)]$$

$$+ \text{ higher order terms} \tag{6.38}$$

Notice that the term containing the first partial derivatives is zero because these derivatives are evaluated at the mode. Let

$$Q_{gh} = - \left. \frac{\partial^2 \ln\{P[X_o(j),Y(1),\ldots,Y(k)]\}}{\partial x_g(j)\partial x_h(j)} \right|_{\substack{x_g(j) = \hat{x}_g(j) \\ \\ x_h(j) = \hat{x}_h(j)}}$$

Then for small $|x_g(j) - \hat{x}_g(j)|$

$$P[X_o(j),Y(1),\ldots,Y(k)] \cong P[\hat{X}_o(j),Y(1),\ldots,Y(k)]$$

$$\cdot \exp\left\{ - \frac{1}{2} \sum_{g,h} Q_{gh}[x_g(j) - \hat{x}_g(j)][x_h(j) - \hat{x}_h(j)] \right\}$$
$$\tag{6.39}$$

Or, over a small region about the mode, $P[X_o(j)|Y(1),\ldots,Y(k)]$ follows the gaussian density.[*]

By picking the elements of K_W arbitrarily small and using Tchebysheff's inequality, $|x_g(j) - \hat{x}_g(j)|$ may be bound as small as desired with any probability less than one. So, with high probability, Eq. (6.39) will describe the joint density. Under these conditions, linear estimation is very nearly optimum.

---

[*]This treatment was suggested by the method used in Ref. 14.

VII. CONCLUSION

A. SUMMARY

A solution has been given for the problem of filtering a stationary process with unknown parameters when the parameters come from some infinite set. It has been shown that the optimum estimator weights are averages of functions of the parameters where the averages are taken with respect to the parameter space conditioned on the observations. It is seen that, for stationary or near-stationary processes, the parameters enter into the solution in the form of simple functions of the signal and noise spectrum. Optimum methods for measuring the spectral averages were given.

The theory of state-variable estimation has been extended to include nonlinear estimation for nongaussian inputs. First, it was shown that, even in steady state, the state variables are nongaussian for nongaussian inputs if the discrete system is stable (a bounded output if the input is bounded). Thus, in general, linear estimation is optimum only for gaussian inputs. The estimation approach used was to find the mode of the state-variable density conditioned on the observations, in which the key problem was to obtain the density in a convenient form. The Markov property of the state variables was used to simplify this very complicated density, and surface-searching methods were used to find the maximum. It has been shown that near optimum (linear or nonlinear) estimates may be made of the state of many dynamical systems using only a short sequence of observations, and the length of this sequence may be related directly to the rate of decay of initial conditions in the dynamical system.

A new approach to linear estimation of state variables has also been presented. It shows the close relationship between the theory of pattern recognition in a random environment and state-variable estimation. The theory is a straightforward extension of Abramson and Braverman's learning theory [Ref. 7]. Their mean and covariance matrix equations are almost identical to the filtering equations presented here. The theory of estimation for the case where the observations are taken through a second dynamical system was presented and a suitable solution was given.

The linear estimation theory was used to simplify the procedure for detecting a gaussian signal in gaussian noise. The optimum detector contains an operator that gives the best estimate of all sample values during the detection interval. Ordinarily, a matrix of order equal to the number of observation sample values must be inverted. It was shown that with state-variable estimation these signal estimates could be made by inverting small matrices of a fixed order independent from the number of observations. A near-optimum, time-invariant detector was derived.

B. SUGGESTIONS FOR FUTURE WORK

Cox [Ref. 6] has investigated state-variable estimation when the system is nonlinear and the inputs are gaussian. By contrast, the present study discusses estimation when the system is linear and the inputs are nongaussian. An obvious area for further work is estimation when the system is nonlinear and the inputs are nongaussian. An approach similar to Chapter VI might be fruitful.

A problem that is considered by the author to be one of the more important unsolved practical problems is the detection of a signal when the signal parameters are unknown and drawn from an infinite set. This is the type of signal that is encountered in practically all space communications. The receiver is not interested in the exact signal or signal parameter, but only in knowing if the signal is present. Intuitively, it is felt that a form of a near optimum detector would correlate the signal and noise with the output of the adaptive filter of Chapter V. This remains to be proven, however. This work should then be extended to include two signals transmitting binary information, where the parameters are drawn from two infinite but not necessarily disjoint sets. An example would be frequency-shift keying when the doppler shift was greater than the keying shift.

APPENDIX A. DERIVATION OF THE SMOOTHING EQUATIONS

This appendix gives the details of the derivation of the smoothing equations. It is desired to find the solution of

$$\text{grad}_{X(1)} \left\{ \ln P[X(1)|Y(1),\ldots,Y(k-1)] \right.$$

$$\left. + \text{grad}_{X(1)} \ln P[Y(k)|X(1),Y(1),\ldots,Y(k-1)] \right\} = 0$$

$$(3.17c)$$

The density $P[Y(k)|X(1),Y(1),\ldots,Y(k-1)]$ may be specified by its mean and covariance matrix. From Eq. (2.3a) it is seen that $E[X(k)|X(1), Y(1),\ldots,Y(k-1)]$ is equal to $\Phi E[X(k-1)|X(1),Y(1),\ldots,Y(k-1)]$, since $W(k-1)$ has zero mean. From Eq. (2.3c) $E[Y(k)|X(1),Y(1),\ldots,Y(k-1)]$ is equal to $H\Phi E[X(k-1)|X(1),Y(k),\ldots,Y(k-1)]$ since $U(k)$ also has zero mean. The solution to the filtering problem, Eq. (3.8), may be used to find $E[X(k-1)|X(1),Y(1),\ldots,Y(k-1)]$ with $X(1)$ substituted for $\bar{X}(1)$ as the initial condition in Eq. (3.14), i.e., for $k = 2$,

$$\hat{X}(2) = \left[H^t K^{-1}_{W(2)} H + K^{-1}_{X(2)} \right]^{-1} \left[H^t K^{-1}_{W(2)} Y(2) + K^{-1}_{X(2)} \Phi X(1) \right] \quad (A.1)$$

where

$$K_{X(2)} = \Gamma K_{U(1)} \Gamma^t \quad (A.2)$$

The matrix $K_{X(k)}$ was given by Eq. (3.6) for $k > 2$.

To simplify notation, Eq. (3.8) can be written as

$$\hat{X}(k-1) = A_{k-1} Y(k-1) + B_{k-1} \hat{X}(k-2) \quad (A.3)$$

Then, iterating on Eq. (A.3), with (A.1) as a start, write the relations

$$\hat{X}(2) = A_2 Y(2) + B_2 X(1)$$

$$\hat{X}(3) = A_3 Y(3) + B_3 A_2 Y(2) + B_3 B_2 X(1)$$

$$\hat{X}(4) = A_4 Y(4) + B_4 A_3 Y(3) + B_4 B_3 A_2 Y(2) + B_4 B_3 B_2 X(1)$$

$$= X(4)\Big|_{X(1)=0} + B_4 B_3 B_2 X(1)$$

More generally,

$$\hat{X}(k-1) = \hat{X}(k-1)\Big|_{X(1)=0} + C_{k-1} X(1) \qquad (A.4)$$

where

$$C_{k-1} \overset{\triangle}{=} B_{k-1} B_{k-2} \cdots B_2 \qquad (A.5)$$

and where $\hat{X}(k-1)\big|_{X(1)=0}$ is the estimate of $X(k-1)$ from the filter equations when $X(1)$ is zero.

It can easily be seen that the covariance matrix of $Y(k)$ given all observations to time $k-1$ is

$$K_{Y(k)}(k|k-1) = H\left[\Phi K_{\hat{X}(k-1)}\Phi^t + \Gamma K_{U(k-1)}\Gamma^t\right]H^t + K_{W(k)} \qquad (A.6)$$

Then

$$\text{grad } \ln\{P[Y(k)|X(1),Y(1),\ldots,Y(k-1)]\}$$

$$= -2C_{k-1}^t \Phi^t H^t K_{Y(k)}^{-1}(k|k-1)\ [Y(k) - H\Phi\hat{X}(k-1)]$$

This gradient is added to the gradient of $\ln\{P[X(1)|Y(1),\ldots,Y(k-1)]\}$ and the sum is set equal to zero to get a solution for $\hat{X}(1)$. The equation to be solved has the form

$$\mathcal{G}_{k-1}\hat{X}(1) - J_{k-1} + C_{k-1}^t \Phi^t H^t K_{Y(k)}^{-1}(k|k-1)\ H\Phi C_{k-1}\hat{X}(1)$$

$$+ C_{k-1}^t \Phi^t H^t K_{Y(k)}^{-1}(k|k-1)\ H\Phi\hat{X}(k-1)\Big|_{X(1)=0}$$

$$- C_{k-1}^t \Phi^t H^t K_{Y(k)}^{-1}(k|k-1)\ Y(k) = 0 \qquad (A.7)$$

from which

$$\hat{X}(1) = \left[C_{k-1}^t \phi^t H^t K_{Y(k)}^{-1}(k|k-1) H\phi C_{k-1} + \mathcal{G}_{k-1} \right]^{-1}$$

$$\cdot \left[C_{k-1}^t \phi^t H^t K_{Y(k)}^{-1}(k|k-1) \left\{ Y(k) - H\phi\hat{X}(k-1) \Big|_{X(1)=0} \right\} + J_{k-1} \right] \quad \text{(A.8)}$$

The quantity $\mathcal{G}_{k-1}$ is the part of the gradient of the exponent of

$$P[X(1)] \ldots P[Y(k-1)|X(1),Y(1),\ldots,Y(k-2)]$$

that is multiplied by $X(1)$, and J_{k-1} is the part of the same gradient that is not multiplied by $X(1)$. For example,

$$\mathcal{G}_2 = K_{X(1)}^{-1} + H^t K_{W(1)}^{-1} H + \phi^t H^t \left[K_{W(2)} + H\Gamma K_{U(2)} \Gamma^t H^t \right]^{-1} H\phi \quad \text{(A.9)}$$

and

$$J_2 = \phi^t H^t \left[K_{W(2)} + H\Gamma K_{U(2)} \Gamma^t H^t \right]^{-1} Y(2) + H^t K_{W(1)}^{-1} + K_{X(1)}^{-1} \bar{X}(1) \quad \text{(A.10)}$$

More generally,

$$\mathcal{G}_k = \left[C_{k-1}^t \phi^t H^t K_{Y(k)}^{-1}(k|k-1) H\phi C_{k-1} + \mathcal{G}_{k-1} \right] \quad \text{(A.11)}$$

and

$$J_k = \left[C_{k-1}^t \phi^t H^t K_{Y(k)}^{-1}(k|k-1) \left\{ Y(k) - H\phi\hat{X}(k-1) \Big|_{X(1)=0} \right\} + J_{k-1} \right] \quad \text{(A.12)}$$

If data before $Y(1)$ are used, Eq. (3.15) can be modified to

$$P[X(1)|\hat{X}(0),Y(1),Y(2),\ldots,Y(k),Y(0),\ldots,Y(-n)]$$

$$= P[X(1)|\hat{X}(0),Y(0),\ldots,Y(-n)] \, P[Y(1)|X(1),\hat{X}(0),Y(0),\ldots,Y(-n)] \, P[\hat{X}(0),Y(0),\ldots,Y(-n)]$$

$$\cdot \frac{P[Y(2)|X(1),Y(1),Y(0),\ldots,Y(-n)] \ldots P[Y(k)|X(1),Y(1),Y(2),\ldots,Y(k-1),Y(0),\ldots,Y(-n)]}{P[\hat{X}(0),Y(1),\ldots,Y(k),Y(0),\ldots,Y(-n)]}$$

$$\text{(A.13)}$$

Note that

$$P[X(1)|\hat{X}(0),Y(0),\ldots,Y(-n)] = P[X(1)|\hat{X}(0)] \qquad (A.14)$$

$$P[Y(1)|X(1),\hat{X}(0),Y(0),\ldots,Y(-n)] = P[Y(1)|X(1)] \qquad (A.15)$$

Also,

$$P[Y(k)|X(1),Y(1),Y(2),\ldots,Y(k-1),Y(0),\ldots,Y(-n)]$$

has mean $H\Phi X[k-1|Y(-n),\ldots,Y(k); X(-n-1) = 0]$, where $\hat{X}(k-1)$ is the mean given in Eqs. (A.1), (A.2), and (A.3) [this can be seen by the same reasoning as that used to derive Eqs. (A.1), (A.2), and (A.3)]. The covariance matrix was given by Eq. (A.6). Thus, the only term of the gradient in Eq. (3.17) that will be changed will be the term due to $P[X(1)|\hat{X}(0)]$.

APPENDIX B. THE CHANGE IN THE DENSITY OF THE CORRELATOR OUTPUT

This appendix presents the change with $\overline{\Delta e^2(i)}$ of the correlation output density of the near-optimum detector of Chapter IV, and gives a method for answering the very important question of how small to make $\overline{\Delta e^2}$. The law of large numbers shows that the correlator density is very nearly gaussian for large n so that only the change in the mean and variance need be discussed.

Before proceeding further, we illustrate the well-known fact that error is uncorrelated with and hence independent of the estimate. From Eq. (4.31),

$$\overline{e^2(i)} = R_x(0) - [A^{(i)}]^t R_{x+w} A^{(i)} \tag{B.1}$$

Since

$$x_1(i) = \hat{x}_1(i) + e(i) \tag{B.2}$$

then

$$\overline{e^2(i)} = \overline{x_1^2(i)} - \overline{\hat{x}_1^2(i)} - 2E[\hat{x}_1(i) \, e(i)] \tag{B.3}$$

However,

$$\overline{\hat{x}_1^2(i)} = [A^{(i)}]^t R_{x+w} A^{(i)} \tag{B.4}$$

and therefore

$$E[\hat{x}_1(i) \, e(i)] = 0 \tag{B.5}$$

1. Change of the Correlator Mean with $\overline{\Delta e^2(i)}$

The output of the correlator is

$$C(n) = \frac{1}{(n - 2m)} \sum_{j=m}^{n-m} \hat{x}_1(j)\, Y(j) \tag{B.6}$$

Using Eq. (B.5), one obtains

$$E[\hat{x}_1(i)\, Y(i)] = E[\hat{x}_1(i)\{\hat{x}_1(i) + e(i) + W(i)\}]$$

$$= E[\hat{x}_1^2(i)] + E[W(i)\, \hat{x}_1(i)]$$

$$\tag{B.7}$$

Since the signal and noise are independent,

$$E[W(i)\, \hat{x}_1(i)] = E\left[W(i) \sum_j a_{ij}\{W(j) + x_1(j)\}\right]$$

$$= a_{ii}\, E[W^2(i)] \tag{B.8}$$

From the first theorem in Chapter IV it is possible to bound the change in $|a_{ii}|$ as m increases. It is also known that the change in $E[\hat{x}_1^2(i)]$ will be less than $\overline{\Delta e^2}$. Then the change in the mean of $C(n)$ can be bounded.

2. The Change in Variance with $\overline{\Delta e^2(i)}$

Write

$$E\left[\left\{[\Delta A^{(i)}]^t\, Y_n\right\}^2\right] = [\Delta A^{(i)}]^t\, R_{x+w}\, \Delta A^{(i)}$$

$$= \sum_k \sum_j \Delta a_{ik}\, \Delta a_{ij}\, R_{x+w}(k-j) \tag{B.9}$$

Since $\|R_{x+w}(t)\|$ is maximum when t equals zero, the sum is maximized for a constant $\|\Delta A\|$ by letting ΔA be different from zero at only one place, i.e.,

$$\Delta a_{ij} = \pm \|\Delta A^{(i)}\| \tag{B.10}$$

Then

$$\max E\left[\left\{[\Delta A^{(i)}]^t Y_n\right\}^2\right] \leq \|\Delta A^{(i)}\|^2 R_{x+w}(0) \tag{B.11}$$

Now $\|[\Delta A^{(i)}]^t Y_n\|$ can be specified to be less than some small fraction of $\sigma^2_{\hat{x}_1(i)}$ with a desired probability. For example, if

$$\frac{\|\Delta A^{(i)}\|}{\|\vec{A}^{(i)}\|} < \frac{1}{40} \tag{B.12}$$

the probability that $\|[\Delta A^{(i)}]^t Y_n\| > 1/10 \, \sigma_{\hat{x}_1(i)}$ is less than 10^{-5}.

If $\|\Delta A^t Y_n\|$ is specified to be less than $1/a \, \sigma_{\hat{x}_1(i)}$ the magnitude of the change in $E[x_1^2(i) \, Y^2(i)]$ as $n \to \infty$ will be less than

$$\left| \int_{Y_\infty} \left(\frac{1}{a} \sigma_{\hat{x}_1(i)} + \hat{x}_1(i)\Big|_{n=\infty}\right)^2 Y^2(i) \, dF(Y_\infty) - E\left[\hat{x}_1^2(i)\Big|_{n=\infty} Y^2(i)\right] \right|$$

$$= \left| \int_{Y_\infty} \left[\frac{1}{a^2} \sigma^2_{\hat{x}_1(i)} Y^2(i) + \frac{2}{a} \sigma_{\hat{x}_1(i)} \hat{x}_1(i)\Big|_{n=\infty} Y^2(i)\right] dF(Y_\infty) \right|$$

$$\cong \frac{2\sigma_{\hat{x}_1(i)}}{a} \int_{Y_\infty} Y^2(i) \sum_j a_{ij} \hat{x}_1(i)\Big|_{n=\infty} Y^2(i) \, dF(Y_\infty) \tag{B.13}$$

- 75 -

Using Eqs. (B.3) and (B.4) and noting that $a_{ii} \lesssim 1$, the magnitude of the difference is less than

$$\frac{2}{a} \sigma^4_{\hat{x}_1(i)} + \frac{2}{a} \sigma_{\hat{x}_1(i)} \sigma^3_{w(i)} \qquad \text{(B.14)}$$

A similar expansion shows that the magnitude of $E[\hat{x}_1(i)Y(i) \cdot \hat{x}_1(j)Y(j)]$ is changed by less than

$$\frac{2}{u} \sigma^2_{\hat{x}_1(i)} \qquad \text{(B.15)}$$

The techniques of Roe and White [Ref. 15] may be used to determine the correlator variance, and Eqs. (B.14) and (B.15) may be used to bound the total change. The two methods may be compared to decide whether the ratio

$$\frac{\|\Delta A^{(i)}\|}{\vec{A}^{(i)}}$$

and thus $\overline{\Delta e^2}$, is small enough. These methods will probably give an unduly pessimistic answer. Usually, a ratio of 1/10 will be sufficient.

APPENDIX C. DERIVATION OF THE FORM OF THE NEAR-OPTIMUM FILTER FOR CONTINUOUS PARAMETER PROCESSES

This appendix contains the derivation of the equations for an adaptive filter which may be synthesized from many narrowband filters. This filter will have a particularly simple adaptive procedure. Consider the system of Fig. 8. The G_i are ideal nonoverlapping filters Δf in bandwidth which completely fill the bandpass $-B$ to B; the G_{o_i} are the optimum smoothing filters following the G_i. Appendix D shows that the error for a steady-state smoothing filter is

$$\overline{e^2} = R_s(0) - \int_0^\infty f^2(t)\, dt$$

$$= R_s(0) - \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} \left| \mathcal{L}\mathcal{F}^{-1}\left\{\frac{S_{ss}}{S_{\overline{ii}}}\right\} \right|^2 ds \qquad (C.1)$$

where

$$f(t) = \mathcal{F}^{-1}\,\mathcal{L}\mathcal{F}^{-1}\left\{\frac{S_{ss}}{S_{\overline{ii}}}\right\} \qquad (C.2)$$

and $R_s(\tau)$ is the autocorrelation function of the signal; $\mathcal{L}\mathcal{F}^{-1}$ is the Laplace transform of the inverse Fourier transform; $S_{ss}(f)$ is the signal power spectrum, and $S_{ii}(f)$ is the spectrum of the signal plus the noise. Assume that $\sum_i G_i$ equals an ideal filter of bandwidth B (equal to unity over the frequency range of interest).

Now, examine the error in the ith channel of Fig. 8. The signal and noise spectra are

$$S_{ss_{(i)}} = S_{ss}\, |G_i|^2 \qquad (C.3)$$

$$S_{nn_{(i)}} = S_{nn}\, |G_i|^2 \qquad (C.4)$$

Assume no correlation between signal and noise. Then

$$S_{ii_{(i)}} = |G_i|^2 \, [S_{ss} + S_{nn}] = |G_i|^2 \, S_{ii} \qquad\qquad (C.5)$$

$$S_{i\bar{i}_{(i)}} \overset{\triangle}{=} G_i^* \, S_{i\bar{i}} \qquad\qquad (C.6)$$

$$S_{ii_{(i)}}^+ \overset{\triangle}{=} G_i \, S_{ii}^+ \qquad\qquad (C.7)$$

and

$$\mathcal{LF}^{-1}\left\{ \frac{S_{ss_{(i)}}}{S_{i\bar{i}_{(i)}}} \right\} = \mathcal{LF}^{-1}\left\{ \frac{S_{ss}}{S_{i\bar{i}}} \, G_i \right\} \qquad\qquad (C.8)$$

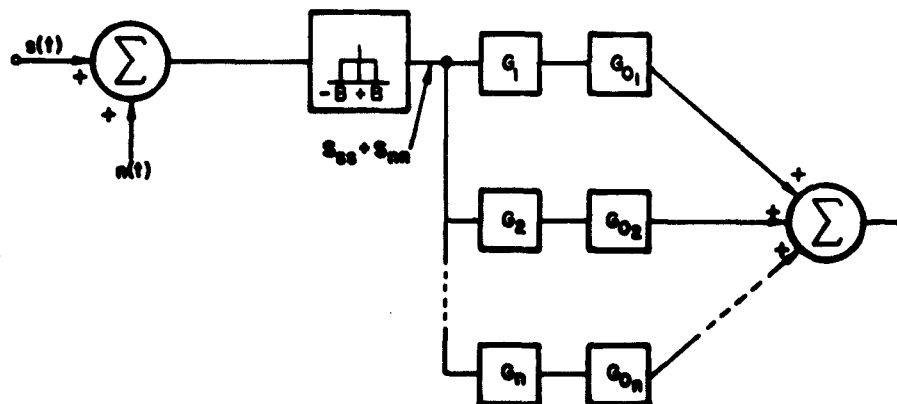FIG. 8.   BLOCK DIAGRAM OF NARROWBAND PARALLEL FILTER SYSTEM.

Since there is no crosscorrelation between channels, the total error for all channels is

$$\overline{e_T^2} = \int_{-\infty}^{\infty} S_{ss} \sum_{i=1}^{n} |G_i|^2 \, df - \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} \sum_{i=1}^{n} \left| \mathcal{LF}^{-1}\left\{ \frac{S_{ss}}{S_{i\bar{i}}} \, G_i \right\} \right|^2 \, ds \quad (C.9)$$

The first integral on the right is

$$\int_{-\infty}^{\infty} S_{ss} \sum_{i=1}^{n} |G_i|^2 \, df = R_s(0) \qquad (C.10)$$

Now evaluate the second integral:

$$\left| \sum_{i=1}^{n} \mathcal{LF}^{-1}\left\{ \frac{S_{ss}}{S_{i\bar{i}}} G_i \right\} \right|^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} \mathcal{LF}^{-1}\left\{ \frac{S_{ss}}{S_{i\bar{i}}} G_i \right\} \left[\mathcal{LF}^{-1}\left\{ \frac{S_{ss}}{S_{i\bar{i}}} G_j \right\} \right]^* \qquad (C.11)$$

Let Δf be small so that $S_{ss}(f)$ and $S_{nn}(f)$ are constant over the interval. Then

$$\mathcal{LF}^{-1}\left\{ \frac{S_{ss}}{S_{i\bar{i}}} G_i \right\} = \frac{S_o(f_i)}{[N_o(f_i) + S_o(f_i)]^{\frac{1}{2}}} \qquad \mathcal{LF}^{-1}(G_i) = \frac{S_o(f_i)}{[N_o(f_i) + S_o(f_i)]^{\frac{1}{2}}} G_i$$

$$(C.12)$$

where $S_o(f_i)$ and $N_o(f_i)$ are the spectral densities in the interval. Also,

$$\left[\mathcal{LF}^{-1}\left\{ \frac{S_{ss}}{S_{i\bar{i}}} G_j \right\} \right]^* = \frac{S_o(f_j)}{[S_o(f_j) + N_o(f_j)]^{\frac{1}{2}}} G_j^* \qquad (C.13)$$

Note that if $G_j^*(f) \neq 0$, then $G_i(f) = 0$; and if $G_i(f) \neq 0$, then $G_j^*(f) = 0$.

$$\frac{1}{2\pi j} \int_{-j\infty}^{j\infty} \mathcal{LF}^{-1}\left\{ \frac{S_{ss}}{S_{i\bar{i}}} G_i \right\} \left[\mathcal{LF}^{-1}\left\{ \frac{S_{ss}}{S_{i\bar{i}}} G_j \right\} \right]^* ds = 0, \qquad i \neq j \qquad (C.14)$$

Assume that S_{ss}, S_{ii}, and G_i can be expressed as the ratio of polynomials (they can always be approximated as closely as desired by such a ratio). Then the partial fraction expansion can be made as follows:

$$\frac{S_{ss}}{S_{i\bar{i}}} G_i = \sum_{\ell} \frac{K_{\ell i}}{S + a_{\ell i}} + \sum_{m} \frac{K_{mi}}{S - b_{mi}} \qquad (C.15)$$

where $a_{\ell i} \geqq 0$ and $b_{mi} \geqq 0$. The right-hand sum has an inverse Fourier transform that is zero for $t \geqq 0$. So,

$$\sum_{i=1}^{n} \mathscr{L}\mathscr{F}^{-1}\left\{\frac{S_{ss}}{S_{i\bar{i}}} G_i\right\} = \sum_{i=1}^{n} \sum_{\ell} \frac{K_{\ell i}}{S + a_{\ell i}} \qquad (C.16)$$

Thus,

$$\mathscr{L}\mathscr{F}^{-1}\left\{\frac{S_{ss}}{S_{i\bar{i}}}\right\} = \mathscr{L}\mathscr{F}^{-1}\left\{\frac{S_{ss}}{S_{i\bar{i}}} \sum_{i=1}^{n} G_i\right\} = \mathscr{L}\mathscr{F}^{-1}\left\{\sum_{i=1}^{n} \frac{S_{ss}}{S_{i\bar{i}}} G_i\right\} = \sum_{i=1}^{n} \mathscr{L}\mathscr{F}^{-1}\left\{\frac{S_{ss}}{S_{i\bar{i}}} G_i\right\}$$

$$(C.17)$$

Using Eqs. (C.11), (C.14), and (C.17), one can see that the second integral of Eq. (C.9) is
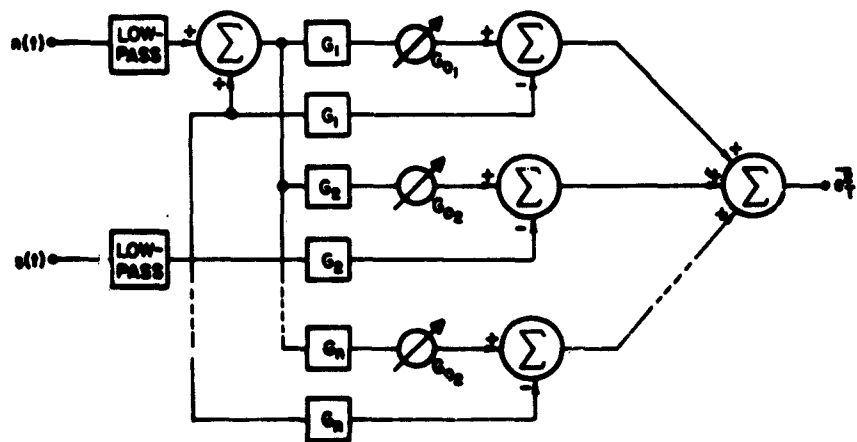
$$\frac{1}{2\pi j} \int_{-j\infty}^{j\infty} \sum_{i=1}^{n} \left| \mathscr{L}\mathscr{F}^{-1}\left\{\frac{S_{ss}}{S_{i\bar{i}}} G_i\right\} \right|^2 ds = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} \left| \mathscr{L}\mathscr{F}^{-1}\left\{\frac{S_{ss}}{S_{i\bar{i}}}\right\} \right|^2 ds \qquad (C.18)$$

Also note that

$$G_{o_i} = \frac{1}{G_i [N_o(f_i) + S_o(f_i)]^{\frac{1}{2}}} \mathscr{L}\mathscr{F}^{-1}\left[\frac{S_o(f_i) G_i}{[N_o(f_i) + S_o(f_i)]^{\frac{1}{2}}} \right] = \frac{S_o(f_i)}{N_o(f_i) + S_o(f_i)}$$

$$(C.19)$$

Figure 9 and Eqs. (C.9), (C.18), and (C.19) show that if the outputs of the parallel filters are weighted by

$$\frac{S_o(f_i)}{N_o(f_i) + S_o(f_i)} = \frac{S_i}{N_i + S_i}$$

a. Block diagram for $\overline{e_T^2}$
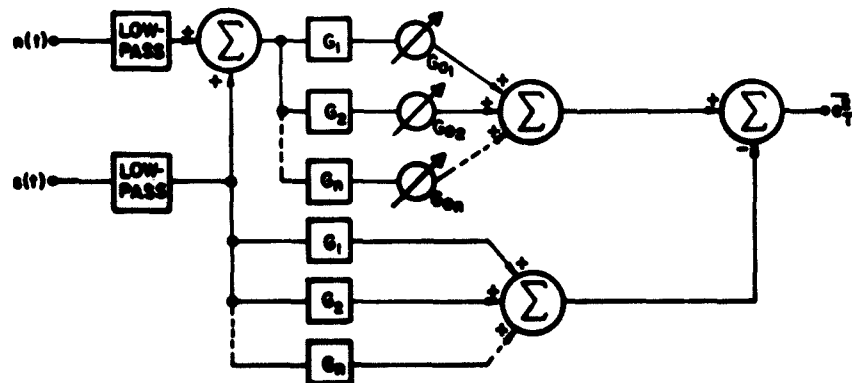
b. Block-diagram manipulation of Fig. 9c

A-32380

c. Equivalent of Fig. 9b

FIG. 9. DIAGRAM FOR FINDING THE ERROR OF THE PARALLEL FILTERS.

where S_i and N_i are signal and noise powers [see Eq. (C.20)], and summed, the resulting transfer function is Wiener optimum. If a number of such narrowband filters can b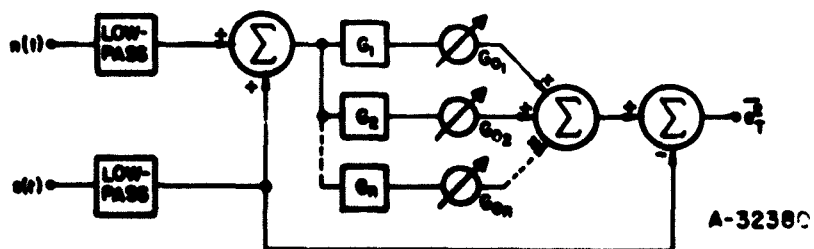e constructed, and if estimates of the signal and noise powers from the narrowband filters are made, then weighting potentiometers with zero-to-one ranges can be adjusted to give an adaptive Wiener filter.

Shortly, it will be shown how Fig. 10 may be modified to use nonideal filters. But, first, the effects of a misadjustment of the potentiometers will be investigated. A reasonable design objective would be an adaptive filter that was some specified decibels worse than optimum. For example, one might want the ratio of the mean-squared error in each channel to the average signal power per channel to be "within 10 percent of that obtainable if the signal spectrum were known exactly." Let K be the nonideal potentiometer setting. The mean-squared error with this setting is

$$\overline{e^2} = \int_{-\infty}^{\infty} [S_{ss} + K^2 S_{ii} - 2S_{ss}K]|G_i|^2 \, df \qquad (C.20)$$

Again assuming nearly constant spectra,

$$\Delta f \ S_{ii} = S + N$$

$$\Delta f \ S_{ss} = S$$

(The ith channel is being considered, and for convenience the subscript i will be dropped.) If $\overline{e_o^2}$ is the minimum mean-squared error, define:

$$D_o = \overline{e^2} - \overline{e_o^2} = \left[K^2(S + N) - 2SK + \frac{S^2}{S + N} \right] \qquad (C.21)$$

Then, solving Eq. (C.21) for K gives

$$K = \frac{S - [D_o(S + N)]^{\frac{1}{2}}}{S + N} \qquad (C.22)$$

and

$$\left| K - G_{o_i} \right| = \frac{[D_o(S + N)]^{1/2}}{S + N}$$

$$= \begin{cases} \left| \frac{\hat{S}}{S + N} - \frac{\hat{S} - \epsilon}{\hat{S} - \epsilon + N} \right| \leq \frac{\epsilon}{\hat{S} + N} & \epsilon > 0 \\[12pt] \left| \frac{\hat{S}}{\hat{S} + N} - \frac{\hat{S} + |\epsilon|}{\hat{S} + |\epsilon| + N} \right| \leq \frac{|\epsilon|}{\hat{S} + |\epsilon| + N} & \epsilon < 0 \end{cases} \qquad (C.23)$$

where ϵ is the measurement error of S, and N is assumed to be known exactly a priori (this is reasonable since there is usually an arbitrarily long time to determine N). Then

$$D_o \leq \frac{\epsilon^2}{S + N} \qquad (C.24)$$
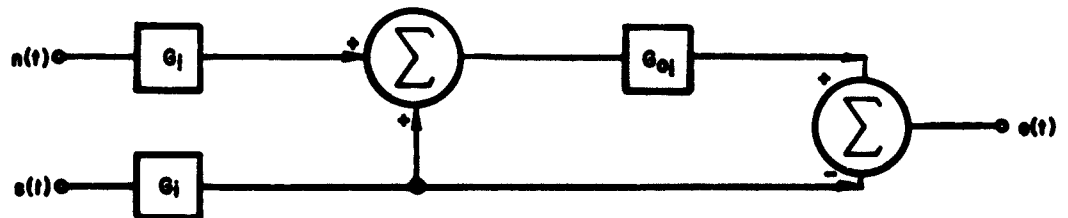
a. Error spectrum due to signal and noise
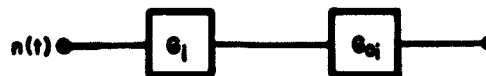
b. Error spectrum due to noise

c. Error spectrum due to signal

FIG. 10. THE BLOCK DIAGRAMS FOR FINDING THE ERROR SPECTRUM OUT OF THE ith CHANNEL.

The foregoing analysis has assumed ideal bandpass filters, i.e., no crosscorrelation between filters. In practice, of course, it is not possible to construct ideal filters. Assume that the ith filter is not ideal but is sharp and overlaps only its adjacent neighbors. Then the additional cross-power error due to the ith channel is (see Fig. 10)
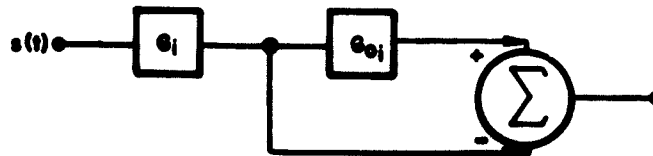
$$\overline{e^2_{c_i}} = \int_{-\infty}^{\infty} R\ell \left\{ G_i \left[G_{o_i} - 1 \right] \left[G^*_{i-1} \left(G_{o_{i-1}} - 1 \right) + G^*_{i+1} \left(G_{o_{i+1}} - 1 \right) \right] \right\} S_o \, df$$

$$+ \int_{-\infty}^{\infty} R\ell \left\{ G_i G_{o_i} \left[G^*_{i-1} G_{o_{i-1}} + G^*_{i+1} G_{o_{i+1}} \right] \right\} N_o \, df$$

$$\leq 4 \int_0^{\infty} |G_i| \left[|G_{i-1}| + |G_{i+1}| \right] \max \left\{ S_o, N_o \right\} df \qquad (C.25)$$

Equation (C.25) provides a very simple means for determining how sharp the filter cutoff should be.

There are now three conditions on the n narrowband filters that will allow $\overline{e^2_T}$ to approach $\overline{e^2}$ arbitrarily closely:

1. The G_i overlap an arbitrarily small amount,

2. $\displaystyle\sum_{i=1}^{n} G_i(j\omega) = Ke^{-j\omega\tau}$, where K is a constant and τ is the delay

    through each G_i over the frequency range of interest.

3. The bandwidth of G_i is arbitrarily small, or the spectra into G_i are nearly constant.

A synthesis procedure to meet these conditions will now be described. If each G_i is to cover the frequency range

$$\frac{B}{n} (i - 1) \quad \text{to} \quad \frac{B}{n} i$$

the G_i can be formed by taking the difference between two lowpass filters with identical linear phase shift or delay; i.e.,

$$G_i = e^{-s\tau} (H_i - H_{i-1})$$
(C.26)

where $e^{-\tau} H_i(s)$ passes from zero to $(B/n)i$, and H_i is the transfer function of a zero-shift sharp-cutoff filter. Obviously,

$$\sum_{i=1}^{n} G_i = e^{-s\tau} H_i$$
(C.27)

thereby fulfilling condition 2.

If H_i is ideal, its impulse response is

$$h_i(t) = \frac{\sin\left(2\pi t \frac{B}{n} i\right)}{\pi t}$$
(C.28)

The transfer function $e^{-s\tau} H_i(s)$ can be made realizable by truncating $h_i(t)$ at $\pm\tau$. Then the impulse response of the nonideal $H_i(s)$ is

$$h_i(t) = \frac{\sin\left(2\pi t \frac{B}{n} i\right)}{\pi t} b(t)$$
(C.29)

where

$$b(t) = \begin{cases} 1 & \cdot \tau \leq t \leq \tau \\ 0 & \text{elsewhere} \end{cases}$$

The transform of $b(t)$ is a sinc function; taking the transform of $h_i(t)$, it is seen that $H_i(s)$ is the convolution of the sinc function with the ideal lowpass filter (see Fig. 11). Then, in the vicinity of $2\pi(B/n)i$,

$$H_i \left(2\pi \frac{B}{n} i + \frac{a}{\tau}\right) = \frac{1}{\tau} \int_{a}^{\infty} \frac{\sin x}{x} \, dx$$
(C.30)

- 85 -

The integral is tabulated, giving a simple method for determining the delay needed for a specified sharpness of cutoff.

A bandlimited filter with an impulse response 2τ long may be constructed with a tapped delay line and a lowpass filter [Ref. 16] as shown in Fig. 12. The setting of the potentiometer on the jth tap, a_{ij}, when synthesizing $e^{-s\tau} H_i(s)$ is

$$a'_{ij} = \frac{\sin\left[\frac{2\pi B i}{n} (jT - \tau)\right]}{\pi(jT - \tau)} , \qquad T \leq \frac{1}{2B} \qquad \text{(C.31)}$$

A practical form of the adaptive filter is shown in Fig. 13.

$$a_j^{(i)} = a'_{ij} - a'_{i(j-1)} \qquad \text{(C.32)}$$

The potentiometers in the ith row correspond to the G_i of Fig. 10, and the signal and noise measurements are made at the output of G_{o_i}. However, if those measurements are made at the output of some other narrowband filter, a great simplification of the adaptive filter is possible. Block diagram substitution will reduce the form of the adaptive filter shown in Fig. 13 to the form shown in Fig. 12, where the jth potentiometer setting is given very simply by

$$a_j = \sum_{i=1}^{n} a_j^{(i)} \frac{S_i}{S_i + N_i} \qquad \text{(C.33)}$$
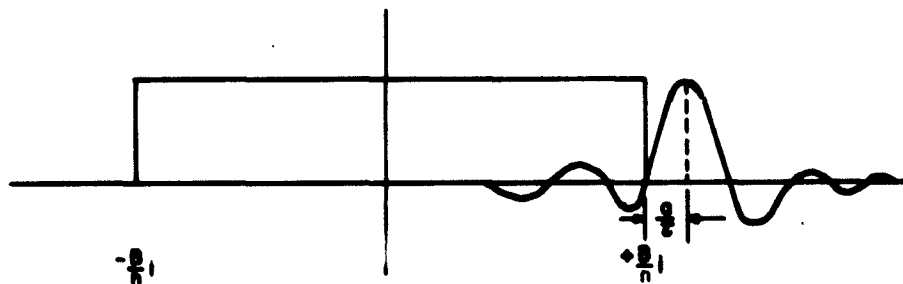
FIG. 11. THE CONVOLUTION.
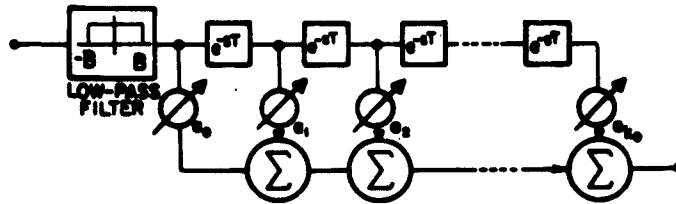
FIG. 12.   SYNTHESIS OF $e^{-s\tau} H_1$ AND OF THE
COMPACT FORM OF THE ADAPTIVE FILTER.
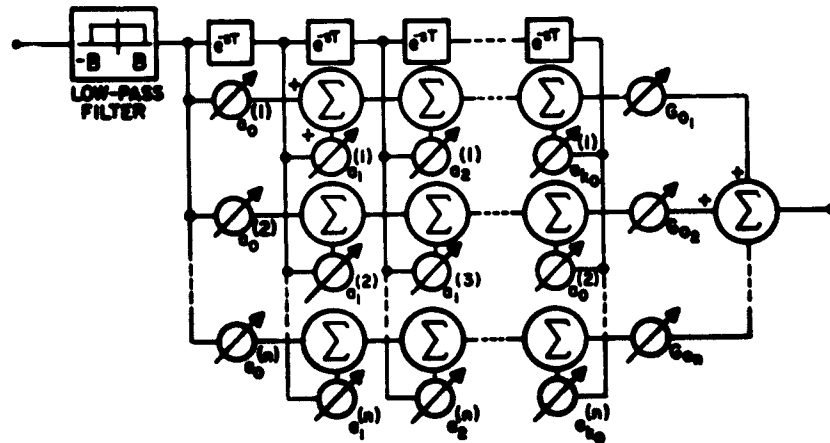
FIG. 13.   A PRACTICAL FORM OF THE ADAPTIVE FILTER.

If the filtering operations are to be performed digitally, one may
go from the delay-line transfer function to numerical techniques in an
almost trivial manner. If f(mT) is the sample value, at t = mT, of
the output of the lowpass filter of Fig. 12, then the sample value at
the output of the sum of the weighted taps is given by

$$c(mT) = \sum_{j=0}^{k_o} a_j f[(m - j)T] \qquad\qquad (C.34)$$

Since c(t) is bandlimited, knowing c(mT) gives all information about
c(t). It is also obvious how the set of difference equations appropriate
to the circuit in Fig. 13 is used.

SEL-64-131

In practice, the spectra will not necessarily be constant over the range of Δf. However, each channel may be considered as the sum of an infinite number of infinitely narrow channels—each with the same potentiometer setting, G_{o_i}. Obviously, this potentiometer setting will not be as good for some of these channels as for others. The equations developed in the estimation part of this appendix may be used to determine Δf under an assumption of the possible slopes of spectra to be encountered. The spectra can vary considerably from a constant value with little harm.

For independent signals and noise, the mean-squared error from an optimum filter is [Ref. 17]

$$\overline{e^2} = \int_{-\infty}^{\infty} df \left\{ |G_d|^2 \, S_{ss} + |G_o|^2 \, S_{ii} - \left[G_o G_d^* + G_d G_o^* \right] S_{ss} \right\} \quad (D.1)$$

where

$$S_{ss} = \text{signal spectrum}$$

$$S_{nn} = \text{noise spectrum}$$

$$S_{ii} = S_{ss} + S_{nn} = S_{ii}^{+} \, S_{i\bar{i}}^{-}$$

$$\mathcal{F}^{-1}\left\{ S_{ii}^{+} \right\} = 0, \quad t < 0$$

$$S_{i\bar{i}}^{-} = \left\{ S_{ii}^{+} \right\}^{*}$$

$$G_d = \text{desired operation}$$

$$G_o = \text{optimum filter given by} \quad G_o(S) = \frac{1}{S_{ii}^{+}} \, \mathcal{L}\mathcal{F}^{-1}\left\{ \frac{S_{ss}}{S_{i\bar{i}}^{-}} \, G_d \right\}$$

Then

$$\overline{e^2} = \int_{-\infty}^{\infty} \left[|G_d|^2 \, S_{ss} + S_{ii} \left| \frac{1}{S_{ii}^{+}} \, \mathcal{L}\mathcal{F}^{-1}\left\{ \frac{S_{ss}}{S_{i\bar{i}}^{-}} \, G_d \right\} \right|^2 - \frac{S_{ss}}{S_{i\bar{i}}^{-}} \, \mathcal{L}\mathcal{F}^{-1}\left\{ \frac{S_{ss}}{S_{i\bar{i}}^{-}} \, G_d \right\} G_d^* \right.$$

$$\left. - \frac{S_{ss}}{S_{i\bar{i}}^{-}} \, \mathcal{L}\mathcal{F}^{-1}\left\{ \frac{S_{ss}}{S_{i\bar{i}}^{-}} \, G_d \right\}^{*} G_d \right] df \quad (D.2)$$

Let

$$f_1(t) = \mathcal{F}^{-1} \mathcal{L} \mathcal{F}^{-1} \left\{ \frac{S_{ss}}{S_{1\bar{1}}} G_d \right\} \qquad (D.3)$$

and

$$f_2(t) = \mathcal{F}^{-1} \left\{ \frac{S_{ss}}{S_{1\bar{1}}} G_d \right\} \qquad (D.4)$$

then

$$f_1(t) = \begin{cases} f_2(t) & t \geqq 0 \\ 0 & t < 0 \end{cases} \qquad (D.5)$$

By Parseval's theorem,

$$\frac{1}{2\pi j} \int_{-j\infty}^{j\infty} \frac{S_{ss}}{S_{1\bar{1}}} G_d \, \mathcal{L}\mathcal{F}^{-1} \left\{ \frac{S_{ss}}{S_{1\bar{1}}} G_d \right\}^* ds = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} \frac{S_{ss}}{S_{1\bar{1}}} G_d^* \, \mathcal{L}\mathcal{F}^{-1} \left\{ \frac{S_{ss}}{S_{1\bar{1}}} G_d \right\} ds$$

$$= \int_{-\infty}^{\infty} f_1(t) \, f_2(t) \, dt = \int_{0}^{\infty} f_1^2(t) \, dt$$

$$= \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} S_{11} \left| \frac{1}{S_{11}^+} \, \mathcal{L}\mathcal{F}^{-1} \left\{ \frac{S_{ss}}{S_{1\bar{1}}} G_d \right\} \right|^2 ds \qquad (D.6)$$

Then

$$\overline{e^2} = \int_{-\infty}^{\infty} |G_d|^2 \, S_{ss} \, df - \int_{0}^{\infty} f_1^2(t) \, dt \qquad (D.7)$$

APPENDIX E.   THE STEADY-STATE MINIMUM-MEAN-SQUARED-ERROR
SAMPLED-DATA FILTER

Assume the filter is to be of the form shown in Fig. 13 and it is desired to adjust the a_i to give minimum mean-squared error. The Z-transform of the tapped-delay-line filter is

$$G(Z) = \sum_0^{k_o} a_n z^{-n} \tag{E.1}$$

The equation for the mean-squared error is well known [Ref. 18] and is

$$\overline{e^2} = R_X(0) + \frac{T}{2\pi j} \oint_{\Gamma_o} \left\{ S_{X+W}(Z) \left[\sum_0^{k_o} a_n z^{-n} \sum_0^{k_o} a_m z^m \right] \right.$$

$$\left. - \sum_{n=0}^{k_o} a_n(z^{-n+\ell} + z^{n-\ell})S_X(Z) \right\} \frac{dZ}{Z} \tag{E.2}$$

The desired operation on the signal is $e^{-\ell T}$. If it is desired to solve for the filter giving the best estimate of the delayed signal, ℓ is a positive integer. If the best predictor is desired, ℓ is a negative integer. The term $R_X(\tau)$ is the signal autocorrelation function; $S_{X+W}(Z)$ is the sampled noise plus signal spectrum; $S_X(Z)$ is the sampled signal spectrum; and the integration is around the unit circle. Uncorrelated signal and noise are assumed.

The optimum setting for a_i may be found by setting the partial derivative of $\overline{e^2}$ with respect to a_i equal to zero and solving for a_i.

$$\frac{\partial \overline{e^2}}{\partial a_i} = \frac{T}{2\pi j} \oint_{\Gamma_o} \frac{dZ}{Z} \left\{ S_{X+W}(Z)z^{-n} \left[\sum_0^{k_o} a_n z^{-n} - \frac{S_X(Z)}{S_{X+W}(Z)} \right] \right\}$$

$$+ \frac{T}{2\pi j} \oint_{\Gamma_o} \frac{dZ}{Z} \left\{ S_{X+W}(Z)z^{-i} \left[\sum_0^{k_o} a_n z^n - \frac{S_X(Z)}{S_{X+W}(Z)} \right] \right\} = 0 \tag{E.3}$$

- 91 -

SEL-64-131

where, for simplicity, $\ell = 0$. Parseval's theorem for discrete systems is

$$\frac{1}{2\pi j} \oint_{\Gamma_o} F_1(Z)\, F_2(Z^{-1})\, \frac{dZ}{Z} = \frac{1}{2\pi j} \oint_{\Gamma_o} F_1(Z^{-1})\, F_2(Z)\, \frac{dZ}{Z}$$

$$= \sum_{-\infty}^{\infty} f_1(nT)\, f_2(nT) \qquad (E.4)$$

Then a solution to Eq. (E.3) is

$$\frac{1}{2\pi j} \oint_{\Gamma_o} \frac{dZ}{Z} \left\{ S_{X+W}(Z) Z^i \left[\sum_{0}^{k_o} a_n\, Z^{-n} - \frac{S_X(Z)}{S_{X+W}(Z)} \right] \right\}$$

$$= \frac{T}{2\pi j} \oint_{\Gamma_o} \frac{dZ}{Z} S_{X+W}(Z) \sum_{0}^{k_o} a_n\, Z^{i-n} - \frac{T}{2\pi j} \oint_{\Gamma_o} S_X(Z)\, Z^i\, \frac{dZ}{Z} = 0$$

$$\qquad (E.5)$$

Again, using Parseval's theorem,

$$\frac{1}{2\pi j} \oint_{\Gamma_o} \frac{dZ}{Z} S_{X+W}(Z) \sum_{0}^{k_o} a_n\, Z^{i-n} = \sum_{n=0}^{k_o} \frac{R_{X+W}[(n-i)T] a_n}{T} \qquad (E.6)$$

and taking the inverse

$$\frac{1}{2\pi j} \oint_{\Gamma_o} S_X(Z)\, Z^i\, \frac{dZ}{Z} = \frac{R_X(iT)}{T} \qquad (E.7)$$

we have (k_o+1) equations of the form

$$\sum_{n=0}^{k_o} a_n R_{x+w}[(n-i)T] = R_x(iT) \qquad (E.8)$$

which can be expressed in matrix form as

$$
\begin{bmatrix}
R_{x+w}(0) & R_{x+w}(T) & R_{x+w}(2T) & \cdots & R_{x+w}(k_oT) \\
R_{x+w}(T) & R_{x+w}(0) & R_{x+w}(T) & & \\
R_{x+w}(2T) & R_{x+w}(T) & R_{x+w}(0) & & \vdots \\
\vdots & \vdots & \vdots & \ddots & \\
R_{x+w}(k_oT) & & \cdots & & R_{x+w}(0)
\end{bmatrix}
\begin{bmatrix}
a_0 \\
a_1 \\
a_2 \\
\vdots \\
a_{k_o}
\end{bmatrix}
=
\begin{bmatrix}
R(0) \\
R_x(T) \\
R_x(2T) \\
\vdots \\
R_x(k_oT)
\end{bmatrix}
$$

$$(E.9)$$

REFERENCES

1. R. E. Kalman and R. W. Koepcke, "Optimal Synthesis of Linear Sampling Control Systems Using Generalized Performance Indexes," ASME Trans., 80, 1958, pp. 1820-26.

2. H. E. Rauch, "Linear Estimation of Sampled Stochastic Processes with Random Parameters," Rept. SEL-62-058 (TR No. 2108-1), Stanford Electronics Laboratories, Stanford, Calif., Apr 1962.

3. B. Widrow, "Adaptive Sampled Data Systems," International Federation of Automatic Control Conv. Rec., Moscow, 1960, p. BP-1-4.

4. D. Gabor, W. P. L. Wilby, and R. Woodcock, "A Universal Non-Linear Filter Predictor and Simulator Which Optimizes Itself by a Learning Process," Proc. IEE (London), Part B, 108, Jul 1960, p. 422.

5. D. T. Magill, "Optimal Adaptive Estimation of Sampled Stochastic Processes," Rept. SEL-63-143 (TR No. 6302-3), Stanford Electronics Laboratories, Stanford, Calif., Dec 1963.

6. H. Cox, "Estimation of State Variables for Noisy Dynamical Systems," Sc.D. thesis, Massachusetts Institute of Technology, May 1963.

7. N. Abramson and D. Braverman, "Learning To Recognize Patterns in a Random Environment," Proc. 1962 International Symposium on Information Theory, IRE Trans. on Information Theory, IT-8, 5, Sep 1962, pp. 58-63.

8. T. Kailath, "Correlation Detection of Signals Perturbed by a Random Channel," IRE Trans. on Information Theory, Jun 1960, pp. 361-366.

9. N. Abramson and J. Farison, "On Applied Decision Theory," Rept. SEL-62-095 (TR No. 2005-2), Stanford Electronics Laboratories, Stanford, Calif., Sep 1962, p. 13.

10. W. B. Davenport, Jr. and W. L. Root, An Introduction to the Theory of Random Signals and Noise, McGraw-Hill Book Co., Inc., New York, 1958.

11. N. Abramson and J. Farison, "On Statistical Communication Theory," Rept. SEL-62-078 (TR No. 2005-1), Stanford Electronics Laboratories, Stanford, Calif., Aug 1962, p. 70.

12. M. Loéve, Probability Theory, D. Van Nostrand Co., Inc., Princeton, N.J., 1960.

13. W. L. Deemer and I. Olken, "Jacobians of Matrix Transformations," Biometrika, 38, 1951, pp. 345-366.

14. L. A. Wainstein and V. P. Zubakov, Extraction of Signals from Noise, Sec. 52, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1962.

15. G. M. Roe and G. M. White, "Probability Density Functions for Correlators with Noisy Reference Signals," IRE Trans. on Information Theory, IT-7, Jan 1961, pp. 13-18.

16. G. L. Turin, "An Introduction to Matched Filters," IRE Trans. on Information Theory, IT-6, Jun 1960, pp. 311-329.

17. O. J. M. Smith, Feedback Control Systems, McGraw-Hill Book Co., Inc., New York, 1958, p. 174.

18. J. R. Ragazzini and G. F. Franklin, Sampled-Data Control Systems, McGraw-Hill Book Co., Inc., New York, 1958, Chapter 14.

GOVERNMENT

USAEL
Ft. Monmouth, N.J.
1 Attn: Dr. H. Jacobs,
AMSEL-RD/SL-PF

Procurement Data Division
USAS Equipment Support Agency
Ft. Monmouth, N.J.
1 Attn: Mr. M. Rosenfeld

Commanding General, USAEL
Ft. Monmouth, N.J.
5 AMSEL-RD/SL-SC, Bldg. 42
1 TDC, Evans Signal Lab Area

Commanding Officer, ERDL
Ft. Belvoir, Va.
1 Attn: Tech. Doc. Ctr.

Commanding Officer
Frankford Arsenal
Bridge and Tacony St.
Philadelphia 37, Pa.
1 Attn: Library Br., 02.0, Bldg. 40

Ballistics Research Lab
Aberdeen Proving Ground, Md.
2 Attn: V. W. Richard, BML

Chief of Naval Research
Navy Dept.
Washington 25, D.C.
2 Attn: Code 427
1   Code 420
1   Code 463

U.S. Army Electr. Labs.
Mt. View Office
P.O. Box 205
1 Mt. View, Calif.

Commanding Officer
ONR Branch Office
1000 Geary St.
1 San Francisco 9, Calif.

Chief Scientist
ONR Branch Office
1030 E. Green St.
1 Pasadena, Calif.

Office of Naval Research
Branch Office Chicago
230 N. Michigan Ave.
1 Chicago 1, Ill.

Commanding Officer
ONR Branch Office
207 W. 24th St.
New York 11, N.Y.
1 Attn: Dr. I. Rowe

U.S. Naval Applied Science Lab.
Tech. Library
Bldg. 291, Code 9832
Naval Base
1 Brooklyn, N Y. 11251

Chief Bureau of Ships
Navy Dept.
Washington 25, D.C.
1 Attn: Code 691A1
1   Code 686
1   Code 607 NTDb
1   Code 687D
1   Code 732, A. E. Smith
1   Code 681A

Officer in Charge, ONR
Navy 100, Bx. 39, Fleet P.O.
16 New York, N.Y.

U.S. Naval Research Lab
Washington 25, D.C.
6 Attn: Code 2000
1   5240
1   5430
1   5200
1   5300
1   5400
1   5266, G. Abraham
1   2027
1   5260
1   6430

Chief, Bureau of Naval Weapons
Navy Dept.
Washington 25, D.C.
1 Attn: RAAV-6
1   RUDC-1
2   RREN-3
1   RAAV-44

Chief of Naval Operations
Navy Dept.
Washington 25, D.C.
1 Attn: Code Op 945Y

Director, Naval Electronics Lab
1 San Diego 52, Calif.

USN Post Graduate School
1 Monterey, Calif.
1 Attn: Tech. Reports Librarian
1   Prof. Gray, Electronics Dept.
1   Dr. H. Titus

Weapons Systems Test Div.
Naval Air Test Center
Patuxent River, Md.
1 Attn: Library

U.S. Naval Weapons Lab
Dahlgren, Va.
1 Attn: Tech. Library

Naval Ordnance Lab.
Corona, California
1 Attn: Library
1   H. H. Wieder, 423

Commanding Officer (ADL)
USN Air Dev. Ctr.
1 Johnsville, Pa. 18974

Commander
USN Missile Center
Pt. Mugu, Calif.
1 Attn: NO3022

Commanding Officer
U.S. Army Research Office
Box CM, Duke Station
Durham, N.C.
3 Attn: CRD-AA-IP

Commanding General
U.S. Army Materiel Command
Washington 25, D.C.
1 Attn: AMCRD-DE-E
1   AMCRD-RS-PE-E

Department of the Army
Office, Chief of Res. and Dev.
The Pentagon
Washington 25, D.C.
1 Attn: Research Support Div.,
        Rm. 3D442

Office of the Chief of Engineers
Dept. of the Army
Washington 25, D.C.
1 Attn: Chief, Library Br.

Hq., U.S. Air Force
Washington 25, D.C. 20330
1 Attn: AFRSTE

Aeronautical Systems Div.
Wright-Patterson AFB, Ohio
1 Attn: Lt. Col. L. M. Butsch, Jr.
        ASRNE-2
1   ASRNE-2, D. R. Moore
1   ASRNR-32
1   ASRNE-1, Electronic Res. Br.
             Elec. Tech. Lab
1   ASRNCF-2, Electromagnetic
             and Comm. Lab
6   ASRNE-32

Systems Engineering Group (RTD)
Wright-Patterson AFB, Ohio 45433
1 Attn: SEPIR

Commandant
AF Institute of Technology
Wright-Patterson AFB, Ohio
1 Attn: AFIT (Library)

Executive Director
AF Office of Scientific Res.
Washington 25, D.C.
1 Attn: SREE

AFWL (WLL)
2 Kirtland AFB, New Mexico

Director
Air University Library
Maxwell AFB, Ala.
1 Attn: CR-4582

Commander, AF Cambridge Res. Labs
ARDC, L. G. Hanscom Field
Bedford, Mass.
1 Attn: CRTOTT-2, Electronics

Hqs., AF Systems Command
Andrews AFB
Washington 25, D.C.
1 Attn: SCTAE

Asst. Secy. of Defense (R and D)
R and D. Board, Dept. of Defense
Washington 25, D.C.
1 Attn: Tech. Library

Office of Director of Defense
Dept. of Defense
Washington 25, D.C.
1 Attn: Research and Engineering

Institute for Defense Analyses
1666 Connecticut Ave.
Washington 9, D.C.
1 Attn: W. E. Bradley

Defense Communications Agency
Dept. of Defense
Washington 25, D.C.
1 Attn: Code 121A, Tech. Library

Advisory Group on Electron Devices
346 Broadway, 8th Floor East
New York 13, N.Y.
2 Attn: H. Sullivan

Advisory Group on Reliability of
Electronic Equipment
Office Asst. Secy. of Defense
The Pentagon
1 Washington 25, D.C.

Commanding Officer
Diamond Ordnance Fuze Labs
Washington 25, D.C.
2 Attn: ORDTL 930, Dr. R.T. Young

Diamond Ordnance Fuze Lab.
U.S. Ordnance Corps.
Washington 25, D.C.
1 Attn: ORDTL-450-638
        Mr. R. H. Comyn

U. S. Dept. of Commerce
National Bureau of Standards
Boulder Labs
Central Radio Propagation Lab.
1 Boulder, Colorado
2 Attn: Miss J.V. Lincoln, Chief
        RWSS

NSF, Engineering Section
1 Washington, D.C.

Information Retrieval Section
Federal Aviation Agency
Washington, D.C.
1 Attn: MS-112, Library Branch

DDC
Cameron Station
Alexandria 4, Va.
30 Attn: TISIA

U.S. Coast Guard
1300 E. Street, N.W.
Washington 25, D.C.
1 Attn: EEE Station 5-5

Office of Technical Services
Dept. of Commerce
1 Washington 25, D.C.

Director
National Security Agency
Fort George G. Meade, Md.
1 Attn: R42

NASA, Goddard Space Flight Center
Greenbelt, Md.
1 Attn: Code 611, Dr. G. H. Ludwig
1   Chief, Data Systems Divisions

NASA
Office of Adv. Res. and Tech.
Federal Office Bldg. 10-B
600 Independence Ave.
Washington, D.C.
1 Attn: Mr. Paul Johnson

Chief, U.S. Army Security Agency
Arlington Hall Station
2 Arlington 12, Virginia

SCHOOLS

*U of Aberdeen
Dept. of Natural Philosophy
Marischal College
Aberdeen, Scotland
1 Attn: Mr. R.V. Jones

U of Arizona
EE Dept.
Tucson, Ariz.
1 Attn: R.L. Walker
1   D.J. Hamilton

*U of British Columbia
Vancouver 8, Canada
1 Attn: Dr. A.C. Soudack

California Institute of Technology
Pasadena, Calif.
1 Attn: Prof. R.W. Gould
1   D. Braverman, EE Dept.

California Institute of Technology
4800 Oak Grove Drive
Pasadena 3, Calif.
1 Attn: Library, Jet Propulsion Lab.

U. of California
Berkeley 4, Calif.
1 Attn: Prof. R.M. Saunders, EE Dept.
        Dr. R.K. Wakerling,
        Radiation Lab. Info. Div.
        Bldg. 30, Rm. 101

U of California
Los Angeles 24, Calif.
1 Attn: C.T. Leondes, Prof. of
        Engineering, Engineering Dept.
1   R.S. Elliott, Electromagnetics
        Div., College of Engineering

U of California, San Diego
School of Science and Engineering
La Jolla, Calif.
1 Attn: Physics Dept.

Carnegie Institute of Technology
Schenley Park
Pittsburg 13, Pa.
1 Attn: Dr. E.M. Williams, EE Dept.

Case Institute of Technology
Engineering Design Center
Cleveland 6, Ohio
1 Attn: Dr. J. B. Neswick, Director

Cornell U
Cognitive Systems Research Program
Ithaca, N.Y.
1 Attn: F. Rosenblatt, Hollister Hall

Thayer School of Engr.
Dartmouth College
Hanover, New Hampshire
1 Attn: John W. Strohbehn
        Asst. Professor

Drexel Institute of Technology
Philadelphia 4, Pa.
1 Attn: F. B. Haynes, EE Dept.

U of Florida
Engineering Bldg., Rm. 336
Gainsville, Fla.
1 Attn: M.J. Wiggins, EE Dept.

Georgia Institute of Technology
Atlanta 13, Ga.
1 Attn: Mrs. J.H. Crosland, Librarian
1   F. Dixon, Engr. Experiment
        Station

Harvard U
Pierce Hall
Cambridge 38, Mass.
1 Attn: Dean H. Brooks, Div. of Engr.
        and Applied Physics, Rm. 217
2   E. Farkas, Librarian, Rm. 303A,
        Tech. Reports Collection

U of Hawaii
Honolulu 14, Hawaii
1 Attn: Asst. Prof. K. Najita, EE Dept.

U of Illinois
Urbana, Ill.
1 Attn: P.D. Coleman, EE Res. Lab.
1   W. Perkins, EE Res. Lab.
1   A. Albert, Tech.Ed.,EE Res.Lab.
1   Library Serials Dept.
1   Prof. D. Alpert, Coordinated
        Sci. Lab.

State University of Iowa
Dept. of Electrical Engineering
Iowa City, Iowa
1 Attn: Prof. Donald L. Epley

*Instituto de Pesquisas da Marinha
Ministerio da Marinha
Rio de Janeiro
Estado da Guanabara, Brazil
1 Attn: Roberto B. da Costa

Johns Hopkins U
Charles and 34th St.
Baltimore 18, Md.
1 Attn: Librarian, Carlyle Barton Lab.

Johns Hopkins U
8621 Georgia Ave.
Silver Spring, Md.
1 Attn: N. H. Choksy
1   Mr. A.W. Nagy, Applied
        Physics Lab.

Linfield Research Institute
McMinnville, Ore.
1 Attn: G. N. Hickok, Director

Marquette University
College of Engineering
1515 W. Wisconsin Ave.
Milwaukee 3, Wis.
1 Attn: A.C. Moeller, EE Dept.

M I T
Cambridge 39, Mass.
1 Attn: Res. Lab. of Elec., Doc. Rm.
1   Miss A. Sils, Libn.Rm. 4-244,
        LIR
1   Mr. J.E. Ward, Elec.Sys.Lab.

M I T
Lincoln Laboratory
P.O. Box 73
1 Attn: Lexington 73, Mass.
1   Navy Representative
1   Dr. W.I. Wells
1   Kenneth L. Jordon, Jr.

U of Michigan
Ann Arbor, Mich.
1 Attn: Dir., Cooley Elec. Labs.,
        N. Campus
1   Dr. J.E. Rowe, Elec. Phys. Lab.
1   Comm. Sci. Lab., 180 Frieze Bldg.

U of Michigan
Institute of Science and Technology
P.O. Box 618
Ann Arbor, Mich.
1 Attn: Tech. Documents Service
1   W. Wolfe--IRIA-- "

U of Minnesota
Institute of Technology
Minneapolis 14, Minn.
1 Attn: Prof. A. Van der Ziel,
        EE Dept.

U of Nevada
College of Engineering
Reno, Nevada
1 Attn: Dr. R.A. Manhart, EE Dept.

Northeastern U
The Dodge Library
Boston 15, Mass.
1 Attn: Joyce E. Lunde, Librarian

Northwestern U
2422 Oakton St.
Evanston, Ill.
1 Attn: W.S. Toth Aerial
        Measurements Lab.

U of Notre Dame
South Bend, Ind.
1 Attn: E. Henry, EE Dept.

Ohio State U
2024 Niel Ave.
Columbus 10, Ohio
1 Attn: Prof. E.M. Boone, EE Dept.

Oregon State U
Corvallis, Ore.
1 Attn: H.J. Oorthuys, EE Dept.

Polytechnic Institute
333 Jay St.
Brooklyn, N.Y.
1 Attn: L. Shaw, EE Dept.

* No AF or Classified Reports

Polytechnic Institute of Brooklyn
Graduate Center, Route 110
Farmingdale, N.Y.
1 Attn: Librarian

Purdue U
Lafayette, Ind.
1 Attn: Library, EE Dept.

Rensselaer Polytechnic Institute
School of Engineering
Troy, N.Y.
1 Attn: Library, Serials Dept.
1   Kenneth E. Mortenson

*U of Saskatchewan
College of Engineering
Saskatoon, Canada
1 Attn: Prof. R.E. Ludwig

Syracuse U
Syracuse 10, N.Y.
1 Attn: EE Dept.

*Uppsala U
Institute of Physics
Uppsala, Sweden
1 Attn: Dr. P. A. Tove

U of Toledo
Dept. of Electr. Engr.
Toledo 6, Ohio
1 Attn: James B. Farison
        Asst. Prof.

U of Utah
Salt Lake City, Utah
1 Attn: R.W. Grow, EE Dept.

U of Virginia
Charlottesville, Va.
1 Attn: J.C. Wyllie, Alderman
        Library

U of Washington
Seattle 5, Wash.
1 Attn: A. E. Harrison, EE Dept.

Worchester Polytechnic Inst.
Worchester, Mass.
1 Attn: Dr. M.H. Newell

Yale U
New Haven, Conn.
1 Attn: Sloane Physics Lab.
1   EE Dept.
1   Dunham Lab., Engr. Library

INDUSTRIES

Avco Corp.
Res. Lab.
2385 Revere Beach Parkway
Everett 49, Mass.
1 Attn: Dr. Gurdon Abell

Argonne National Lab.
9700 South Cass
Argonne, Ill.
1 Attn: Dr. O.C. Simpson

Admiral Corp.
3800 Cortland St.
Chicago 47, Ill.
1 Attn: E.N. Roberston, Librarian

Airborne Instruments Lab.
Comac Road
Deer Park, Long Island, N Y.
1 Attn: J. Dyer, Vice-Pres. and
        Tech. Dir.

Amperex Corp.
230 Duffy Ave.
Hicksville, Long Island, N.Y.
1 Attn: Proj. Engineer, S. Barbasso

*No AF or Classified Reports.

Autonetics
Div. of North American Aviation, Inc.
9150 E. Imperial Highway
Downey, Calif.
1 Attn: Tech. Library 3040-3

Bell Telephone Labs.
Murray Hill Lab.
Murray Hill, N.J.
1 Attn: Dr. J.R. Pierce
1   Dr. S. Darlington
1   Mr. A.J. Grossman

Bell Telephone Labs., Inc.
Technical Information Library
Whippany, N.J.
1 Attn: Tech. Repts. Librn.,
        Whippany Lab.

The Boeing Company
Mail Stop MS-1331-ORG. 1-8000
Seattle 24, Washington
1 Attn: Dr. Ervin J. Nalos

*Central Electronics Engineering
Research Institute
Pilani, Rajasthan, India
1 Attn: Om P. Gandhi - Via: ONR/London

Columbia Radiation Lab.
538 West 120th St.
1 New York, New York

Convair - San Diego
Div. of General Dynamics Corp.
San Diego 12, California
1 Attn: Engineering Library

Cook Research Labs.
6401 W. Oakton St.
1 Attn: Morton Grove, Ill.

Cornell Aeronautical Labs., Inc.
4455 Genessee
Buffalo 21, N.Y.
1 Attn: Library

Eitel-McCullough, Inc.
301 Industrial Way
San Carlos, Calif.
1 Attn: Research Librarian

Ewan Knight Corp.
East Natick, Mass.
1 Attn: Library

Fairchild Semiconductor Corp.
4001 Junipero Serra Blvd.
Palo Alto, Calif.
1 Attn: Dr. V. H. Grinich

General Electric Co.
Defense Electronics Div., LMED
Cornell University, Ithaca, N.Y.
1 Attn: Library
   Via: Commander, ASD W-P, Ohio,
        ASRNGW D.E. Lewis

General Electric TWT Products Sec.
601 California Ave.
Palo Alto, Calif.
1 Attn: Tech. Library, C.G. Lob

General Electric Co. Res. Lab.
P.O. Box 1088
Schnectady, N.Y.
1 Attn: Dr. P.M. Lewis
1   R.L. Shuey, Mgr. Info.
        Studies Sec.

General Electric Co.
Electronics Park
Bldg. 3, Rm. 143-1
Syracuse, N.Y.
1 Attn: Doc. Library, Y. Burke

Gilfillan Brothers
1815 Venice Blvd.
Los Angeles, Calif.
1 Attn: Engr. Library

The Hallicrafters Co.
5th and Kostner Ave.
1 Attn: Chicago 24, Ill.

Hewlett-Packard Co.
1501 Page Mill Road
1 Attn: Palo Alto, Calif.

Hughes Aircraft
Malibu Beach, Calif.
1 Attn: Mr. Iams

Hughes Aircraft Co.
Florence at Teale St.
Culver City, Calif.
1 Attn: Tech. Doc. Cen., Bldg. 6,
        Rm. C2048

Hughes Aircraft Co.
P.O. Box 278
Newport Beach, Calif.
1 Attn: Library, Semiconductor Div.

IBM, Box 390 Boardman Road
Poughkeepsie, N.Y.
1 Attn: J.C. Logue, Data Systems Div.

IBM Poughkeepsie, N.Y.
1 Attn: Product Dev. Lab.,
        E.M. Davis

IBM ASD and Research Library
Monterey and Cottle Roads
San Jose, Calif.
1 Attn: Miss M. Griffin, Bldg. 025

ITT Federal Labs.
500 Washington Ave.
Nutley 10, N.J.
1 Attn: Mr. E. Mount, Librarian

Laboratory for Electronics, Inc.
1075 Commonwealth Ave.
Boston 15, Mass.
1 Attn: Library

LEL, Inc.
75 Akron St.
Copiague, Long Island, N.Y.
1 Attn: Mr. R.S. Mautner

Lenkurt Electric Co.
San Carlos, Calif.
1 Attn: M.L. Waller, Librarian

Librascope
Div. of General Precision, Inc.
808 Western Ave.
Glendale 1, Calif.
1 Attn: Engr. Library

Lockheed Missiles and Space Div.
P.O. Box 504, Bldg. 524
Sunnyvale, Calif.
1 Attn: Dr. W.M. Harris, Dept. 65-70
1   G. W. Price, Dept. 67-33

Melpar, Inc.
3000 Arlington Blvd.
Falls Church, Va.
1 Attn: Librarian

Microwave Associates, Inc.
Northwest Industrial Park
Burlington, Mass.
1 Attn: K. Mortenson
1   Librarian

Microwave Electronics Corp.
4061 Transport St.
Palo Alto, Calif.
1 Attn: S.F. Kaisel
        M.C. Long

Minneapolis-Honeywell Regulator Co.
1177 Blue Heron Blvd.
Riviera Beach, Fla.
1 Attn: Semiconductor Products Library

The Mitre Corp.
Bedford, Mass.
1 Attn: Library

Monsanto Research Corp.
Station B, Box 8
Dayton 7, Ohio
1 Attn: Mrs. D. Crabtree

Monsanto Chemical Co.
800 N. Linbergh Blvd.
St. Louis 66, Mo.
1 Attn: Mr. E. Orban, Mgr.
        Inorganic Dev.

*Dir., National Physical Lab.
Hilside Road
New Delhi 12, India
1 Attn: S.C. Sharma - Via:
        ONR/London

*Northern Electric Co., Lmtd.
Research and Development Labs.
P.O. Box 3511, Station "C"
Ottawa, Ontario, Canada
1 Attn: J.F. Tatlock

Northronics
Palos Verdes Research Park
6101 Crest Road
Palos Verdes Estates, Calif.
1 Attn: Tech. Info. Center

Pacific Semiconductors, Inc.
14520 So. Aviation Blvd.
Lawndale, Calif.
1 Attn: H.Q. North

Philco Corp.
Tech. Rep. Division
P.O. Box 4730
Philadelphia 34, Pa.
1 Attn: F.R. Sherman, Mgr. Editor

Philco Corp.
Jolly and Union Meeting Roads
Blue Bell, Pa.
1 Attn: C.T. McCoy
1   Dr. J. R. Feldmeier

Polarad Electronics Corp.
43-20 Thirty-Fourth St.
Long Island City 1, N.Y.
1 Attn: A.H. Sonnenschein

Radio Corp. of America
RCA Labs., David Sarnoff Res.Cen.
Princeton, N.J.
2 Attn: Dr. J. Sklansky

RCA Labs., Princeton, N.J.
1 Attn: H. Johnson

RCA, Missile Elec. and
Controls Dept.
Woburn, Mass.
1 Attn: Library

The Rand Corp.
1700 Main St.
Santa Monica, Calif.
1 Attn: Helen J. Waldron, Librarian

Raytheon Manufacturing Co.
Microwave and Power Tube Div.
Burlington, Mass.
1 Attn: Librarian, Spencer Lab

Raytheon Manufacturing Co.
Res. Div., 28 Seyon St.
Waltham, Mass.
1 Attn: Dr. H. Statz
1   Mrs. M. Bennett, Librarian
1   Research Div. Library

Roger White Electron Devices, Inc.
Tall Oaks Road
1 Laurel Hedges, Stamford, Conn

Sandia Corp.
Sandia Base Albuquerque, N.M.
1 Attn: Mrs. B.R. Allen, Librarian

Sperry Rand Corp.
Sperry Electron Tube Div.
Gainesville, Fla.
1 Attn: Librarian

Sperry Gyroscope Co.
Div. of Sperry Rand Corp.
Great Neck, N.Y.
1 Attn: L. Swern (MS3T105)

Sperry Gyroscope Co.
Engineering Library
Mail Station F-7
Great Neck, Long Island, N.Y.
1 Attn: K. Barney, Engr. Dept. Head

Sperry Microwave Electronics
Clearwater, Fla.
1 Attn: J.E. Pippin, Res. Sec. Head

Sylvania Electric Products, Inc.
500 Evelyn Ave.
1 Mt. View, Calif.
1 Attn: Mr. E.O. Ammann

Sylvania Electronics Systems
100 First St.
Waltham 54, Mass.
1 Attn: Librarian, Waltham Labs.
1   Mr. E.E Hollis

Technical Research Group
Route No. 110
1 Melville, New York   11749

Texas Instruments, Inc.
P.O. Box 6015
Dallas 22, Texas
1 Attn: M.E. Chun, Apparatus Div.

Texas Instruments, Inc.
P.O. Box 5012
Dallas, Texas   75222
1 Attn: Tech. Repts. Service, MS-65
2   Semi-Conductor Components
         Library

Texas Instruments, Inc.
Corporate Research and Engineering
Technical Reports Service
P.O. Box 5474
1 Dallas 22, Texas

Tektronix, Inc.
P.O. Box 500
Beaverton, Ore.
4 Attn: Dr. J.F. DeLord, Dir. of
        Research

Varian Associates
611 Hansen Way
Palo Alto, Calif.
1 Attn: Tech. Library

Weitermann Electronics
4549 North 38th St.
1 Milwaukee 9, Wisconsin

Westinghouse Electric Corp.
Friendship International Airport
Box 746, Baltimore 3, Md.
1 Attn: G.R. Kilgore, Mgr. Appl. Res.
        Dept. Baltimore Lab.

Westinghouse Electric Corp.
3 Gateway Center
Pittsburgh 22, Pa.
1 Attn: Dr. G.C. Sziklai

Westinghouse Electric Corp.
P.O. Box 284
Elmira, N.Y.
1 Attn: S.S. King

Zenith Radio Corp
6001 Dickens Ave
Chicago 39, Ill.
1 Attn: J. Markin

*No AF or Classified Reports

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)	2a. REPORT SECURITY CLASSIFICATION
Stanford Electronics Laboratories	UNCLASSIFIED
Stanford University, Stanford, California	2b. GROUP

3. REPORT TITLE

ESTIMATING AND DETECTING THE OUTPUTS OF LINEAR DYNAMICAL SYSTEMS

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical Report

5. AUTHOR(S) (Last name, first name, initial)

C. S. Weaver

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
December 1964	95	18

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
Office of Naval Research	Technical Report No. 6302-7
b. PROJECT NO. Contract Nonr-225(24)	SU-SEL-64-131
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	

10. AVAILABILITY/LIMITATION NOTICES

Qualified requesters may obtain copies from DDC. Foreign announcement and dissemination by DDC limited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	ONR

13. ABSTRACT

This investigation considers three closely related problems: the optimum filtering of stationary or near-stationary random processes with unknown parameters from an infinite parameter set; estimation of the state of a linear discrete dynamical system with nongaussian noisy inputs; and applications of state estimation theory to detection. The form of the optimum filter when the parameters are unknown is found to have weights that are averages of simple functions of the signal and noise spectra averaged over the parameter space. Practical method for implementation are given. The key problem in nonlinear state-variable estimation is obtaining the joint density of the states and the observations in a convenient form. This problem is unsolved, and surface searching is used to find the mode. The number of dimensions of the surface is the same as the order of the dynamical system. A new approach to linear state estimation is given; and this theory is applied to the problem of detecting a gaussian signal in gaussian noise. A time-invariant, near-optimum detector of small dimensions is derived.

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
CONTROL SYSTEMS						
FEEDBACK						
LINEAR SYSTEMS						
DETECTION						
TOPOLOGY						
PROGRAMMING(COMPUTERS)						
MATRIX ALGEBRA						
STATISTICAL ANALYSIS						
NOISE						
STATISTICAL PROCESS						
STOCHASTIC PROCESSES						
SIGNALS						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization (corporate author) issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers (either by the originator or by the sponsor), also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those

imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through

_____ ."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through

_____ ."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through

_____ ."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring (paying for) the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as (TS), (S), (C), or (U).

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.